**Exercises (Chapters 5 – 7) – Solutions**

* 1. To estimate the regression model, go to Quick -> Estimate Equation. In the Equation specification field, write:

price c sqft age

Click OK. Table 1 shows the estimation results. These results indicate the followings: First, the intercept term is estimated to be equal to -41947.70. Thus, for a house with zero surface and zero age the price would be equal to -41947.70. Of course, this number is meaningless (negative prices do not exist!), so we cannot interpret this intercept estimate. Second, the coefficient of the SQFT variable is estimated to be equal to 90.96. If the surface of the house increases (decreases) by 1 square feet, the price of the house will increase (decrease) by $90.96. This coefficient is statistically significant at the 1% level, indicating a statistically significant positive relation between surface and price. Third, the coefficient of the AGE variable is estimated to be equal to -755.04. If the age of the house decreases (increases) by 1 year, the price of the house will increase (decrease) by $755.04. This coefficient is statistically significant at the 1% level, indicating a statistically significant negative relation between age and price. Finally, the coefficient of the regression is 59%, indicating that 59% of the variation in house prices is explained by the model, while the remaining 41% remains unexplained.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 1: OLS estimates regressing PRICE on SQFT and AGE** | | | | |
| Dependent Variable: PRICE | | |  |  |
| Method: Least Squares | | |  |  |
| Sample: 1 1080 | |  |  |  |
| Included observations: 1080 | | |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| C | -41947.70 | 6989.636 | -6.001414 | 0.0000 |
| SQFT | 90.96980 | 2.403100 | 37.85519 | 0.0000 |
| AGE | -755.0414 | 140.8936 | -5.358948 | 0.0000 |
|  |  |  |  |  |
|  |  |  |  |  |
| R-squared | 0.589592 | Mean dependent var | | 154863.2 |
| Adjusted R-squared | 0.588829 | S.D. dependent var | | 122912.8 |
| S.E. of regression | 78814.86 | Akaike info criterion | | 25.39036 |
| Sum squared resid | 6.69E+12 | Schwarz criterion | | 25.40421 |
| Log likelihood | -13707.80 | Hannan-Quinn criter. | | 25.39561 |
| F-statistic | 773.6077 | Durbin-Watson stat | | 1.909534 |
| Prob(F-statistic) | 0.000000 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

* 1. This hypothesis is formalized as follows: The t-statistic of the test is:

The critical value is . As the null hypothesis cannot be rejected at the 1% significance level. Therefore, we cannot accept the hypothesis that having a house one-year older decreases price by 1000 or less.

* 1. The following two graphs show the relation between SQFT, AGE, and the residuals of the previous estimation. The first figure indicates that the relation between SQFT and PRICE may not be linear. For low and high levels of SQFT, the model fitted values are above the true sample values generating positive residuals. For the intermediate levels of SQFT, the model fitted values are below the sample values generating negative residuals. Given, the shape of the plot perhaps a polynomial of order 2 may be more suitable to capture the relation between SQFT and PRICE. The second figure indicates the existence of model uncertainty for younger houses compared to older ones, due to the existence of heteroskedasticity. Alternatively, it may be attributed to the existence of a non-linear relationship between AGE and PRICE.



**Figure 1: Scatter plot between residuals and SQFT**



**Figure 2: Scatter plot between residuals and AGE**

* 1. Table 2 shows the results of the new model with two additional independent variables, and . To estimate this regression model, go to Quick -> Estimate Equation. In the Equation specification field, write:

price c sqft age sqft^2 age^2

* + 1. The marginal effect of on is given by:

So, the estimated marginal effect for the smallest house with is:

The estimated marginal effect for a house with is:

The estimated marginal effect of the largest house with is:

These values indicate that as the size of house increases, the impact of an additional square feet on price increases. The result for the smallest house is unrealistic, and it is due to the model inability to fit this part of the sample.

* + 1. The marginal effect of on is given by:

So, the estimated marginal effect for the oldest house with is:

The estimated marginal effect for a house with is:

The estimated marginal effect for the youngest house with is:

These results indicate that as a house becomes older, the impact of an additional year on price decreases. When the house is new, an extra year has the largest negative effect on price. As the house becomes older this negative impact diminishes. For old houses the impact of on is positive. This is unrealistic, indicating that the quadratic nature of the model cannot fit well on this part of the sample.

* + 1. The marginal effect of on when is given as . The critical value from the t-distribution for a 95% confidence interval is . Therefore, the confidence interval is given as:

where

Thus, the confidence interval is:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 2: OLS estimates regressing PRICE on SQFT and AGE** | | | | |
| Dependent Variable: PRICE | | |  |  |
| Method: Least Squares | | |  |  |
| Sample: 1 1080 | |  |  |  |
| Included observations: 1080 | | |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| C | 170149.6 | 10432.26 | 16.30996 | 0.0000 |
| SQFT | -55.78417 | 6.389443 | -8.730677 | 0.0000 |
| AGE | -2797.788 | 305.1155 | -9.169604 | 0.0000 |
| SQFT^2 | 0.023153 | 0.000964 | 24.01350 | 0.0000 |
| AGE^2 | 30.16033 | 5.071049 | 5.947553 | 0.0000 |
|  |  |  |  |  |
|  |  |  |  |  |
| R-squared | 0.735162 | Mean dependent var | | 154863.2 |
| Adjusted R-squared | 0.734177 | S.D. dependent var | | 122912.8 |
| S.E. of regression | 63371.44 | Akaike info criterion | | 24.95603 |
| Sum squared resid | 4.32E+12 | Schwarz criterion | | 24.97911 |
| Log likelihood | -13471.26 | Hannan-Quinn criter. | | 24.96477 |
| F-statistic | 746.0221 | Durbin-Watson stat | | 1.954354 |
| Prob(F-statistic) | 0.000000 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

* 1. Table 3 shows the results of the new model adding the interaction variable . To estimate this regression model, go to Quick -> Estimate Equation. In the Equation specification field, write:

price c sqft age sqft^2 age^2 sqft\*age

* + 1. The marginal effect of on is given by:

The estimated marginal effect for a house with and is:

The estimated marginal effect for a house with and is:

The estimated marginal effect for a house with and is:

The results are like those reported in part (d). Again, the marginal effect of on increases as the surface of the house increases too.

* + 1. The marginal effect of on is given by:

The estimated marginal effect for a house with and is:

The estimated marginal effect for a house with and is:

The estimated marginal effect for a house with and is:

These results lead to similar conclusions to those reached in part (d). The negative marginal impact of on decreases in magnitude as the house becomes older.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 3: OLS estimates regressing PRICE on SQFT and AGE adding an interaction term** | | | | |
| Dependent Variable: PRICE | | |  |  |
| Method: Least Squares | | |  |  |
| Sample: 1 1080 | |  |  |  |
| Included observations: 1080 | | |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| C | 114597.4 | 12142.85 | 9.437440 | 0.0000 |
| SQFT | -30.72888 | 6.897561 | -4.455035 | 0.0000 |
| AGE | -442.0336 | 410.6123 | -1.076523 | 0.2819 |
| SQFT^2 | 0.022185 | 0.000943 | 23.53737 | 0.0000 |
| AGE^2 | 26.51899 | 4.938587 | 5.369753 | 0.0000 |
| SQFT\*AGE | -0.930621 | 0.112436 | -8.276887 | 0.0000 |
|  |  |  |  |  |
|  |  |  |  |  |
| R-squared | 0.751042 | Mean dependent var | | 154863.2 |
| Adjusted R-squared | 0.749883 | S.D. dependent var | | 122912.8 |
| S.E. of regression | 61470.74 | Akaike info criterion | | 24.89605 |
| Sum squared resid | 4.06E+12 | Schwarz criterion | | 24.92374 |
| Log likelihood | -13437.87 | Hannan-Quinn criter. | | 24.90654 |
| F-statistic | 647.9975 | Durbin-Watson stat | | 2.009009 |
| Prob(F-statistic) | 0.000000 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

* 1. Table 4 reports RESET test results for the two models used in part (d) and (e). For both models the RESET tests reject the null hypothesis at the 1% level, indicating that the models are mispecified either in terms of the selected functional form or they have omitted important variables.

Table 5 reports model selection criteria for the two models used in part (d) and (e). All selection criteria indicate that the model in part (e) provides a better fit to the data examined.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 4: RESET test results, F-statistics and p-values in parenthesis** | | | | |
|  | RESET (numb. fitted values = 1) | | RESET (numb. fitted values = 2) | |
| Model in (d) | 69.45 | (0.00) | 43.34 | (0.00) |
| Model in (e) | 74.27 | (0.00) | 49.87 | (0.00) |

|  |  |  |
| --- | --- | --- |
| **Table 5: Model selection criteria** | | |
|  | Model in (d) | Model in (e) |
| Adjusted | 73.4% | 74.9% |
| AIC | 24.95 | 24.89 |
| BIC | 24.97 | 24.92 |

* 1. To estimate the regression model, go to Quick -> Estimate Equation. In the Equation specification field, write:

log(prod) c log(area) log(labor) log(fert)

Click OK. Table 1 shows the estimation results. These results indicate the followings: First, the intercept term is estimated to be equal to -1.54. Thus, for a farm with the log production is -1.54, indicating that the production is 0.21. Second, the coefficient of the variable is estimated to be equal to 0.36. If the hectares planted increases (decreases) by 1%, the rice farm production will increase (decrease) by 0.36%. This coefficient is statistically significant at the 1% level, indicating a statistically significant positive relation between planted area and production. Third, the coefficient of the variable is estimated to be equal to 0.43. If labor increases (decreases) by 1%, the rice farm production will increase (decrease) by 0.43%. This coefficient is statistically significant at the 1% level, indicating a statistically significant positive relation between labor and production. Fourth, the coefficient of the variable is estimated to be equal to 0.21. If the quantity of fertilizers used increases (decreases) by 1%, the rice farm production will increase (decrease) by 0.21%. This coefficient is statistically significant at the 1% level, indicating a statistically significant positive relation between fertilizers and production. Finally, the coefficient of the regression is 85%, indicating that 85% of the variation in rice farm production is explained by the model, while the remaining 15% remains unexplained.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 1: OLS estimates of a log-log model regressing PROD on AREA, LABOR, and FERT** | | | | |
| Dependent Variable: LOG(PROD) | | |  |  |
| Method: Least Squares | | |  |  |
| Sample: 1 352 | |  |  |  |
| Included observations: 352 | | |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| C | -1.546786 | 0.255654 | -6.050321 | 0.0000 |
| LOG(AREA) | 0.361736 | 0.063968 | 5.654966 | 0.0000 |
| LOG(LABOR) | 0.432848 | 0.066883 | 6.471762 | 0.0000 |
| LOG(FERT) | 0.209502 | 0.038265 | 5.474979 | 0.0000 |
|  |  |  |  |  |
|  |  |  |  |  |
| R-squared | 0.847871 | Mean dependent var | | 1.541307 |
| Adjusted R-squared | 0.846559 | S.D. dependent var | | 0.871600 |
| S.E. of regression | 0.341419 | Akaike info criterion | | 0.699888 |
| Sum squared resid | 40.56536 | Schwarz criterion | | 0.743792 |
| Log likelihood | -119.1802 | Hannan-Quinn criter. | | 0.717360 |
| F-statistic | 646.5085 | Durbin-Watson stat | | 1.776258 |
| Prob(F-statistic) | 0.000000 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

* 1. This hypothesis is formalized as follows: To perform this test in Eviews in the Equation object go to View -> Coefficient Diagnostics -> Wald test. In the new window field write the null hypothesis of your test. To do so, write:

C(2)=C(3)

The t-statistic is -0.58 and the p-value is equal to 55%. Therefore, the null cannot be rejected at all conventional significant levels, indicating that the elasticity of production with respect to land is equal to the elasticity of production with respect to labor. Analytically, we can perform the test using the following t-statistic:

that follows the t-distribution with 348 degrees of freedom. The estimated standard error is equal to:

Therefore, the sample value of the t-statistic is:

The critical value is . Since, the null hypothesis cannot be rejected at the 5% level.

* 1. To perform this test in Eviews in the Equation object go to View -> Coefficient Diagnostics -> Wald test. In the new window field write the null hypothesis of your test. To do so, write:

C(2)+C(3)+C(4)=1

The t-statistic is 0.17 and the p-value is equal to 86%. Therefore, the null cannot be rejected at all conventional significant levels, indicating that the production function exhibits constant return on scales. We can perform the test analytically following the same procedure as in part (b).

* 1. To perform this test in Eviews in the Equation object go to View -> Coefficient Diagnostics -> Wald test. In the new window field write the null hypothesis of your test. To do so, write:

C(2)=C(3), C(2)+C(3)+C(4)=1

The F-statistic is 0.18 and the p-value is equal to 83%. Therefore, the null cannot be rejected at all conventional significant levels, indicating that the elasticity of production with respect to land is equal to the elasticity of production with respect to labor and the production function exhibits constant return on scales.

* + 1. With omitted the elasticity for changes from 0.3617 to 0.4567, and the elasticity for 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 appears to bias the other elasticities upwards, but the omitted variable is not picked up by the RESET.
    2. With omitted the elasticity for changes from 0.3617 to 0.6633, and the elasticity for 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 also appears to bias the other elasticities upwards, but again the omitted variable is not picked up by the RESET.
    3. With omitted the elasticity for changes from 0.2095 to 0.2682, and the elasticity for 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 appears to bias the other elasticities upwards, particularly that for . In this case the omitted variable misspecification has been picked up by the RESET with two extra terms.
  1. Table 1 shows the estimation results. These results indicate the followings: First, the intercept term is estimated to be equal to 0.42. Thus, for a worker with the log wage is 0.42, indicating that the average wage is $1.52. Second, the coefficient of variable is estimated to be equal to 0.11. If the education increases (decreases) by 1 year, average wage rate will increase (decrease) by 11%. This coefficient is statistically significant at the 1% level, indicating a statistically significant positive relation between education and wages. Third, the coefficient of variable is estimated to be equal to 0.012. If expertise increases (decreases) by 1 year, average wage rate will increase (decrease) by 1.2%. This coefficient is statistically significant at the 1% level, indicating a statistically significant positive relation between expertise and wages. Finally, the coefficient of the regression is 26%, indicating that 26% of the variation in log wage rates is explained by the model, while the remaining 74% remains unexplained.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 1: OLS estimates of a log-linear model regressing WAGE on EDUC and EXPER** | | | | |
| Dependent Variable: LOG(WAGE) | | |  |  |
| Method: Least Squares | | |  |  |
| Sample: 1 4733 | |  |  |  |
| Included observations: 4733 | | |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| C | 0.426087 | 0.043006 | 9.907516 | 0.0000 |
| EDUC | 0.113581 | 0.002944 | 38.57749 | 0.0000 |
| EXPER | 0.012057 | 0.000609 | 19.81157 | 0.0000 |
|  |  |  |  |  |
|  |  |  |  |  |
| R-squared | 0.263417 | Mean dependent var | | 2.166535 |
| Adjusted R-squared | 0.263106 | S.D. dependent var | | 0.549766 |
| S.E. of regression | 0.471933 | Akaike info criterion | | 1.336674 |
| Sum squared resid | 1053.469 | Schwarz criterion | | 1.340770 |
| Log likelihood | -3160.240 | Hannan-Quinn criter. | | 1.338114 |
| F-statistic | 845.7734 | Durbin-Watson stat | | 0.499742 |
| Prob(F-statistic) | 0.000000 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

* 1. Table 2 shows the estimation results. These results indicate the followings: First, the intercept term is estimated to be equal to 0.52. Thus, for a male, non-black, and unmarried worker who is not member of the union with the log wage is 0.52, indicating that the average wage is $1.68. Second, the coefficient of variable is estimated to be equal to 0.11. If the education increases (decreases) by 1 year, average wage rate will increase (decrease) by 11%. This coefficient is statistically significant at the 1% level, indicating a statistically significant positive relation between education and wages. Third, the coefficient of variable is estimated to be equal to 0.009. If expertise increases (decreases) by 1 year, average wage rate will increase (decrease) by 0.9%. This coefficient is statistically significant at the 1% level, indicating a statistically significant positive relation between expertise and wages. Fourth, the coefficient of dummy variable is estimated to be equal to -0.23. Thus, all else being equal, an average female worker earn 23% less than a male worker. This coefficient is statistically significant indicating that the gender of the employee impacts his/her salary. Fifth, the coefficient of dummy variable is estimated to be equal to -0.09. Thus, all else being equal, an average black worker earn 9% less than a non-black worker. This coefficient is statistically significant indicating that the race of the employee affects the wage rate. Sixth, the coefficient of dummy variable is estimated to be equal to 0.12. Thus, all else being equal, an average married worker earn 12% more than an unmarried worker. This coefficient is statistically significant indicating that the married status of the employee affects the wage rate. Seventh, the coefficient of dummy variable is estimated to be equal to 0.18. Thus, all else being equal, the average member of the union earn 18% more than a non-member worker. This coefficient is statistically significant indicating that the union membership affects the wage rate. Finally, the coefficient of the regression is 34%, indicating that 34% of the variation in log wage rates is explained by the model, while the remaining 66% remains unexplained.

For a female, married, black individual who is also member of the union the model implies that the average log wage rate is:

For a male, unmarried, white individual who is not member of the union the average log wage rate is:

If we hold constant education and experience the difference is given as:

This difference is estimated to be equal to . Therefore, we conclude that a female, married, black individual who is also member of the union earn 2% less than a male, unmarried, white individual who is not member of the union.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 2: OLS estimates of a log-linear model regressing WAGE on EDUC, EXPER and a set of dummy variables** | | | | |
| Dependent Variable: LOG(WAGE) | | |  |  |
| Method: Least Squares | | |  |  |
| Sample: 1 4733 | |  |  |  |
| Included observations: 4733 | | |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| C | 0.528922 | 0.041214 | 12.83352 | 0.0000 |
| EDUC | 0.110608 | 0.002805 | 39.42658 | 0.0000 |
| EXPER | 0.009666 | 0.000609 | 15.87450 | 0.0000 |
| FEMALE | -0.229078 | 0.013064 | -17.53541 | 0.0000 |
| BLACK | -0.093859 | 0.021960 | -4.274049 | 0.0000 |
| MARRIED | 0.121788 | 0.014035 | 8.677215 | 0.0000 |
| UNION | 0.177644 | 0.017781 | 9.990734 | 0.0000 |
|  |  |  |  |  |
|  |  |  |  |  |
| R-squared | 0.342121 | Mean dependent var | | 2.166535 |
| Adjusted R-squared | 0.341286 | S.D. dependent var | | 0.549766 |
| S.E. of regression | 0.446197 | Akaike info criterion | | 1.225364 |
| Sum squared resid | 940.9061 | Schwarz criterion | | 1.234921 |
| Log likelihood | -2892.824 | Hannan-Quinn criter. | | 1.228723 |
| F-statistic | 409.6155 | Durbin-Watson stat | | 0.642323 |
| Prob(F-statistic) | 0.000000 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

* 1. This hypothesis is formalized as follows: . The sum of squares error of the unrestricted model is . The sum of squares error of the restricted model is . Therefore, the sample valued of the F-statistic is given as:

The critical value is . As the null is rejected, indicating that the four dummy variables are jointly statistically significant at the 5% level.

* 1. To conduct the Chow test, we first need to estimate the unrestricted model:

The Chow test hypothesis is formalized as follows: . The sum of squares error of the unrestricted model is . The sum of squares error of the restricted model is . The sample value of the F-statistic is:

The critical value is . As the null is rejected, indicating that affects the wage equation.

* 1. The marginal effect of on is:

Therefore, for a female individual this marginal effect is estimated to be equal to or 11.7%. For a male individual this marginal effect is estimated to be equal to 10.4%. We conclude that an additional year of education increases the wage of a female individual more than the wage of a male individual. Moreover, this difference is statistically significant (since coefficient is significant at the 5% level).

The marginal effect of on is:

Therefore, for a female individual this marginal effect is estimated to be equal to or 0.84%. For a male individual this marginal effect is estimated to be equal to 1.1%. We conclude that an additional year of expertise increases the wage of a male individual more than the wage of a female individual. Moreover, this difference is statistically significant (since coefficient is significant at the 5% level).

* 1. This hypothesis is formalized as follows: . To perform this test in Eviews in the Equation object go to View -> Coefficient Diagnostics -> Wald test. In the new window field write the null hypothesis of your test. To do so, write:

C(2)+C(8)=0

The t-statistic is 27.82 and the p-value is equal to 0.00. We conclude that the null hypothesis is rejected, indicating that education significantly affects the wage of female individuals.

* 1. Table 3 shows model selection criteria for the three models. All three criteria indicate that the models estimated in part (b) and (e) fit better to the data compared to the model estimated in (a). Between these two models, the BIC prefers model in part (b), while the AIC and the adjusted prefers model in part (e).

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 3: Model selection criteria** | | | |
|  | Model in (a) | Model in (b) | Model in (e) |
| Adjusted | 26.3% | 34.1% | 34.2% |
| AIC | 1.336 | 1.225 | 1.223 |
| BIC | 1.340 | 1.235 | 1.236 |