Exercises (Chapters 5 – 7)

1. The file house.xlsx contains data on 1080 houses sold during 2005. It includes the following variables: the price of the house sold (PRICE), the surface of the house (SQFT, in square feet), and the age of the house (AGE).
   a. Estimate the regression model
   $$PRICE = \beta_1 + \beta_2 SQFT + \beta_3 AGE + \varepsilon$$
   and comment on the results
   b. Test the hypothesis that having a house a year older decreases prices by 1000 or less.
   c. Construct two graphs showing the relation between the two independent variables and the residuals. Also conduct the RESET test. Comment on these results. Is the model correctly specified?
   d. Add the variables $SQFT^2$ and AGE^2 to the model and re-estimate the equation.
      i. Find estimates of the marginal effect $\partial PRICE/\partial SQFT$ for the smallest house in the sample, the largest house in the sample and the house with 2300 square feet. Comment on these values. Are they realistic?
      ii. Find estimates of the marginal effect $\partial PRICE/\partial AGE$ for the oldest house in the sample, the newest house in the sample and the house which is 20 years old. Comment on these values. Are they realistic?
      iii. Find a 95% confidence interval for the marginal effect $\partial PRICE/\partial SQFT$ for a house with 2300 square feet.
   e. Add the interaction variable $SQFT \times AGE$ to the model in part (d) and re-estimate the equation. Report the results. Repeat (i) and (ii) using $SQFT = 2300$ and $AGE = 20$. Compare the results with those of part (d).
   f. Perform a RESET test for the models in parts (d) and (e). Which model, overall, is preferable according to the RESET test results and model selection criteria?

2. The file prod.xlsx includes 352 pooled (cross-sectional and time series) observations from rice farms. It includes tonnes of freshly threshed rice (PROD), hectares planted (AREA), person-days of hired and family labor (LABOR) and kilograms of fertilizer (FERT).
   a. Estimate the following log-log linear regression model (production function):
   $$\ln(PROD) = \beta_1 + \beta_2 \ln(AREA) + \beta_3 \ln(LABOR) + \beta_4 \ln(FERT) + \varepsilon$$
   and comment on the results. Interpret the coefficient estimates.
   b. Using a 5% significance level test the hypothesis that the elasticity of production with respect to land is equal to the elasticity of production with respect to labor.
   c. Using a 10% significance level test the hypothesis that the production function exhibits constant returns to scale, that is, $\beta_2 + \beta_3 + \beta_4 = 1$.
   d. Using a 5% significance level, jointly test the two hypotheses in parts (b) and (c).
   e. Re-estimate the model in part (a) with: (i) FERT omitted, (ii) LABOR omitted, and (iii) AREA omitted. In each case discuss the effect of omitting a variable on the estimates

of the remaining two elasticities. Also, in each case examine if the RESET test with 2 fitted values has picked up the omitted variables.

3. The file wage.xlsx includes data on 4733 individuals. It includes hourly wage rates (WAGE), number of years of education (EDUC), number of years of experience (EXPER) and four dummy variables taking the value of 1 if the individual is a female (FEMALE), a black person (BLACK), married (MARRIED) and member of the union (UNION).
   a. Estimate the log-linear regression model
   $$\ln(WAGE) = \beta_1 + \beta_2 EDUC + \beta_3 EXPER + \varepsilon$$
   and comment on the results. Interpret the coefficient estimates.
   b. Estimate the log-linear regression model in part (a) with the four dummy variables included in it, that is,
   $$\ln(WAGE) = \beta_1 + \beta_2 EDUC + \beta_3 EXPER + \beta_4 FEMALE + \beta_5 BLACK + \beta_6 MARRIED + \beta_7 UNION + \varepsilon$$
   Interpret the coefficient estimates. Is the average wage of a female, married, black individual who is also member of the union larger (or smaller) to the average wage of a male, unmarried, white individual who is not member of the union, holding constant education and experience?
   c. Test at the 5% significance level that the four dummy variables are jointly significant in explaining the cross-sectional variation of wages in the sample.
   d. Conduct the Chow test (at the 5% level) for the difference in the wage equation in part (a) for male and female individuals.
   e. Estimate the following log-linear regression model,
   $$\ln(WAGE) = \beta_1 + \beta_2 EDUC + \beta_3 EXPER + \beta_4 FEMALE + \beta_5 BLACK + \beta_6 MARRIED + \beta_7 UNION + \beta_8 EDUC \times FEMALE + \beta_9 EXPER \times FEMALE + \varepsilon$$
   What is the marginal effect of education and experience for male and female individuals on their wages?
   f. Test at the 5% significance level that education significantly affects female individuals' wages.
   g. Which model in part (a), (b) and (e) is preferable according to model selection criteria?