

## Introductory Econometrics: Computer Problem Sheet 1

I am assuming that you know the basics of Excel. For instance, I assume you know how to load in data and manipulate variables (e.g. takes log, squares, etc. of variables using the formula bar). Regressions can be run using the options which appear when you click on Tools then Data analysis. You will have a tutor in the computer lab to help you with these things. The purpose of this problem set is for you to gain familiarity with running regressions in Excel and interpreting results. I would encourage you to make sure you know the basic commands in your computer lab (if necessary, asking the tutor for help). Then spend some time by yourself experimenting with this data set.

This computer session uses data set HPRICE.XLS (available through the course website at <http://personal.strath.ac.uk/gary.koop/ie.html>) which contains data on N=546 houses sold in Windsor, Canada. Our dependent variable, Y, is the sales price of the house in Canadian dollars. The explanatory variables included in this data set are:

- the lot size of the property (in square feet)
- the number of bedrooms
- the number of bathrooms
- the number of storeys (excluding the basement).
- A dummy variable = 1 if house has a driveway (=0 otherwise)
- A dummy variable = 1 if house has a recreation room (=0 otherwise)
- A dummy variable = 1 if house has a basement (=0 otherwise)
- A dummy variable = 1 if house has gas central heating (=0 otherwise)
- A dummy variable = 1 if house has air conditioning (=0 otherwise)
- The size of garage (number of cars it will hold)
- A dummy variable = 1 if house is in a desirable neighbourhood (=0 otherwise)

In this session, I want you to carry out an empirical analysis of this data set and discuss your empirical results as you would in an empirical project or dissertation *using only the methods associated with ordinary least squares*. The following questions should help structure this process:

- i) Run a regression of the dependent variable on all the explanatory variables. How would you interpret the coefficient estimates? Does the interpretation of the dummy variables differ from the other explanatory variables?
- ii) Are all the explanatory variables statistically significant? Why? If you find insignificant variables, omit them from the regression and repeat part i).
- iii) Is there evidence of multicollinearity?
- iv) What is the  $R^2$ ? How would you interpret this number?
- v) Now consider some extensions of the basic model. Generate a new explanatory variable which is the dummy for “desirable neighbourhood” times the “lot size” variable. Run a regression including all the explanatory variables described above plus this new explanatory variable. How do you interpret the coefficient on this new explanatory variable? Does inclusion of this new variable alter any of your results in parts i) through iv)?
- vi) Generate a new variable which is “lot size” squared and include it in the regression described in part v). How do you interpret the coefficient on this new variable? Does inclusion of this new variable alter any of your results in parts i) through v)?