Introductory Econometrics
Tutorial 10

PART A: This homework is a review of parameter interpretation and serial correlation. After reviewing the lecture slides about dummies and serial correlation, and studying chapter 6.2 and 12 of the textbook, attempt Week 10 Part A quiz on Moodle. You need to submit the quiz before your tutorial and attend the tutorial to obtain 1 point for Week 10 participation.

Part B: This part will be covered in the tutorial. It is still a good idea to attempt these questions before the tutorial.
The purpose of this tutorial is to use dummy variables when using time series data.

1. The data set SeattleElectric2005-6.xlsx contains data on average hourly electricity usage in cold months in Seattle, USA, from 1 Oct 2005 to 31 March 2006. The variables in this data set are:

date	:	date
dow	:	day of the week (1 for Sunday to 7 for Saturday)
pubhol	:	=1 if that day was a public holiday, =0 otherwise
avetemp	:	average daily temperature in degrees Fahrenheit
aveload	:	average hourly electricity load on that day in megawatts

   (a) Use *dow* to create dummy variables *sat* for Saturday (day 7) and *sun* for Sunday (day 1). In addition, create a dummy variable called *wknd* which is 1 if the day is a Saturday or a Sunday, and 0 for any other day.

   (b) The goal is to predict *aveload* in winter months (Seattle, USA, is in the Northern hemisphere) based on temperature and other available information (our assumption is that the weather bureau can produce temperature forecasts that are pretty accurate). Given this goal, have an appropriate look at the data. In particular, investigate differences in average load in different days of the week and look at the scatter plot of *aveload* against *avetemp* to familiarise yourself with the data. Remember that a good analyst does not confine himself or herself to any particular software. If a pivot chart would be better for visualising differences in average load in different days of the week, then use it! If you have learnt another software (for example R, SPSS or STATA) in another unit that you think will be more helpful, use it!

   (c) Estimate a regression of *aveload* on a constant, *avetemp*, *wknd* and the public holiday dummy. We know that electricity usage, like anything else in life, is likely to be related to what the usage was in previous days. As a result, we suspect that part of the electricity load that cannot be explained by temperature is likely to be correlated over time. If that is the case, what consequences would that have for the OLS estimator? In such a case, can we use the model that you estimated to test that there is no difference in the intercept in weekdays and weekends?

   (d) Use visual aids and a formal test to test the hypothesis that errors of this model are white noise against the alternative that they are generated by an AR(7). You should reject the null. Then investigate what kind of AR model would be sufficient for capturing the dynamics of the errors (the t-statistics in your BG auxiliary regression may give you a hint, and the partial autocorrelations of the residuals in the correlogram are informative as well). Re-estimate the model by adding an AR equation for u_t.