Quantitative Techniques:

Laboratory 10

Autocorrelation and heteroskedasticity in simple regression

Overview

In this lab your tasks are:

1. to explore the effect of autocorrelation on
    a. coefficient estimates
    b. estimated standard errors
2. to explore the effect of heteroskedasticity on
    a. coefficient estimates
    b. estimated standard errors

As last week, we have a true model:

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, *i= 1, 2… n*

Last week, we assumed that the error terms had the following properties:

1. normally distributed
2. independent of each other
3. constant variance

This week we are going to explore the effects of relaxing (2) and (3), separately:

- Assumption (2) ONLY is relaxed in Task 2.
- Assumption (3) ONLY is relaxed in Task 3.

You need to know that when (2) is relaxed we have a situation known as **autocorrelation**, and its main effect is to create bias in the standard errors of the regression and the coefficients. As a result *t* tests will reject the null hypothesis too frequently. When (3) is relaxed we have a situation known as **heteroskedasticity**, and like autocorrelation, its main effect is to create bias in the standard errors of the regression and the coefficients. As a result *t* tests will reject the null hypothesis too frequently.

On the Excel menu bar, click on Tools, Add-ins…., then make sure that Analysis ToolPak is ticked. Download *Lab10.xls* from http://www.staff.city.ac.uk/a.banal-estanol/ and save it in your directory.

Task 1: Basic Regression

1. Open the spreadsheet and go to the worksheet named ‘Random variables’;
2. In column B, 100 values of *X* that are uniformly distributed between 10 and 30 have been created;
3. In column C, 100 values of ε_i that are standard normally distributed have been created;
4. In worksheet ‘Regression’, values of Y have been created assuming that in the true model β_0=10 and β_1=2;
5. Obtain estimates for β_0 and β_1. Do they look unbiased?

6. Obtain standard errors for the parameters using the function LINEST(Y, X, 1,1) from Laboratory 8;
7. Calculate a t-test for both parameters;
8. Generate the fitted values of Y, i.e. Yhat = β_0 hat + $\beta_1$hat * X;
9. Generate the residuals, i.e. ei = Y – Yhat.

   You can also change the variance of the errors (but keeping the variances for all observations equal). To change the variance we used the fact that var (kZ) = k^2 var(Z), so we started with a standard normal variable *z* and multiplied it by a constant *c.* so that *u=cz.* The parameter 'c' is in cell C3 in 'Regressions'.

Task 2: Serial Correlation

1. In the worksheet 'Random Variables', autocorrelated errors have been created in Column M;[1]

$\varepsilon_t = \rho * \varepsilon_{t-1} + v_t$

2. Graph the errors in Column M ('Random variables' worksheet) and compare with a similar graph of the errors in Column C. Do they follow any pattern?
3. Now obtain estimates and standard errors for β_0 and β_1 in the worksheet 'Autocorrelation' using the function LINEST(Y, X, 1, 1). Do they look unbiased?
4. Compare the standard errors of the parameters with those you obtained at Task 1;
5. Generate the fitted values of Y, i.e. Yhat = β_0 hat + $\beta_1$hat * X;
6. Generate the residuals, i.e. ei = Y – Yhat;
7. Plot the residuals from this regression and compare your graph with the graph from the regression of Task 1.

Task 3: Heteroskedasticity

1. In the worksheet 'Random Variables', heteroskedastic errors have been created in Column P;
2. Now obtain estimates and standard errors for β_0 and β_1 in the worksheet 'Heteroskedasticity' using the function LINEST(Y, X, 1, 1). Do they look unbiased?
3. Generate the fitted values of Y, i.e. Yhat = β_0 hat + $\beta_1$hat * X;
4. Generate the residuals, i.e. ei = Y – Yhat;
5. Create a 'scatter plot' graph of the fitted values (on the horizontal axis) and the residuals (on the vertical axis);
6. For the residuals in 'Regression', create a scatter plot similar to point 5 and compare.

---

[1] In case you were wondering about the formula in Column M - In order to ensure that ε_t keeps its variance of 1 we have to multiply the value in Column C which represents the v_t component by a constant k since
$\text{var}(\varepsilon_t) = \rho^2 * \text{var}(\varepsilon_{t-1}) + k^2 * \text{var}(v_t)$
where v_t is itself standard normal. So I have to make sure that $k^2 = 1 - \rho^2$.