Introduction to Regression and OLS

Rob Hayward

February 23, 2015

Outline

Modelling

The model

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

Modelling

The model

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

Where

- y_t is the dependent variable

Modelling

The model

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

Where

- y_t is the dependent variable
- α is an intercept or constant

Modelling

The model

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

Where

- y_t is the dependent variable
- α is an intercept or constant
- x_t is the explanatory or independent variable(s)

Modelling

The model

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

Where

- y_t is the dependent variable
- α is an intercept or constant
- x_t is the explanatory or independent variable(s)
- β is the key relationship

Modelling

The model

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

Where

- y_t is the dependent variable
- α is an intercept or constant
- x_t is the explanatory or independent variable(s)
- β is the key relationship
- ε_t is the error that covers omitted variables, measurement error and other stochastic or random elements

Modelling

The model

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

Modelling

The model

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

Where

- y_t is the inflation rate

Modelling

The model

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

Where

- y_t is the inflation rate
- α is an intercept or constant

Modelling

The model

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

Where

- y_t is the inflation rate
- α is an intercept or constant
- x_t is the unemployment rate

Modelling

The model

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

Where

- y_t is the inflation rate
- α is an intercept or constant
- x_t is the unemployment rate
- β is the relationship between the inflation rate and the unemployment rate

Modelling

The model

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

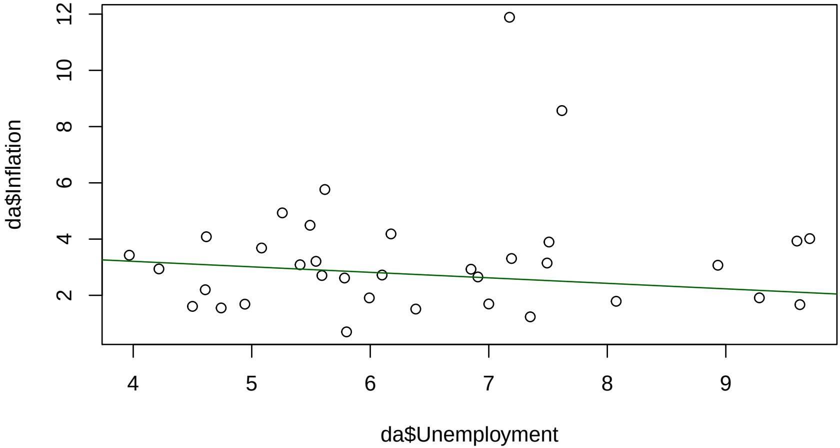
Where

- y_t is the inflation rate
- α is an intercept or constant
- x_t is the unemployment rate
- β is the relationship between the inflation rate and the unemployment rate
- ε_t is all the other factors that affect the inflation rate

Caution!

> *"Essentially all models are wrong, but some are useful"*

Scattergram



R Squared (p. 13)

The total variance of the dependent variable is called the total sum of squares (TSS). This can be split into

- Explained sum of squares or sum of squares of the regression (ESS)

R Squared (p. 13)

The total variance of the dependent variable is called the total sum of squares (TSS). This can be split into

- Explained sum of squares or sum of squares of the regression (ESS)
- Residual sum of squares (RSS)

R Squared (p. 13)

The total variance of the dependent variable is called the total sum of squares (TSS). This can be split into

- Explained sum of squares or sum of squares of the regression (ESS)
- Residual sum of squares (RSS)

R Squared (p. 13)

The total variance of the dependent variable is called the total sum of squares (TSS). This can be split into

- Explained sum of squares or sum of squares of the regression (ESS)
- Residual sum of squares (RSS)

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{RSS}{RSS + ESS}$$

$$R^2 = 1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{(y - \bar{y})'(y - \bar{y})}$$

$$u = \hat{\varepsilon} \tag{1}$$

Adjusted R Squared (p. 13)

The R^2 can be considered a measure of *goodness of fit*. However, the more variables that you add the smaller the R^2. The *Adjusted R Squared* ($\bar{R}^2$) will make a penalty for adding variables.

$$\bar{R}^2 = 1 - (1 - R^2) \times \frac{(T - 1)}{(T - K)} \tag{2}$$

where T is the total number of observations and K is the number of variables.

Coefficient Estimates

Remember that the estimates of the coefficients will depend on the sample

- A different sample will give a different estimate

Coefficient Estimates

Remember that the estimates of the coefficients will depend on the sample

- A different sample will give a different estimate
- We want to know how reliable the estimates will be under different samples

Coefficient Estimates

Remember that the estimates of the coefficients will depend on the sample

- A different sample will give a different estimate
- We want to know how reliable the estimates will be under different samples
- β is a random variable. If OLS (*Gauss-Markov* assumptions hold)

Coefficient Estimates

Remember that the estimates of the coefficients will depend on the sample

- A different sample will give a different estimate
- We want to know how reliable the estimates will be under different samples
- β is a random variable. If OLS (*Gauss-Markov* assumptions hold)
  - $\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\beta_1})$

Coefficient Estimates

Remember that the estimates of the coefficients will depend on the sample

- A different sample will give a different estimate
- We want to know how reliable the estimates will be under different samples
- β is a random variable. If OLS (*Gauss-Markov* assumptions hold)
  - $\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\beta_1})$

Coefficient Estimates

Remember that the estimates of the coefficients will depend on the sample

- A different sample will give a different estimate
- We want to know how reliable the estimates will be under different samples
- β is a random variable. If OLS (*Gauss-Markov* assumptions hold)
  - $\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\beta_1})$

  If we assume a normal distribution we can carry out hypothese tests about coefficients like β_1