ECON2206 Notes

OLS Estimates

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Coefficient of Determination

$$SST = SSE + SSR$$

$$\underbrace{\sum_{i=1}^{n}(y_i - \bar{y})^2}_{Total\ Sum\ of\ Squares} = \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}_{Explained\ Sum\ of\ Squares} + \underbrace{\sum_{i=1}^{n}\hat{u}_i}_{Sum\ of\ Squares\ Residuals}$$

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

Interpretations

Dependent	Independent	Interpretation of β_1
y	x	$\Delta y = \beta_1 \Delta x$ One unit change in 'x' leads to a β_1-unit change in 'y'
y	$\log(x)$	$\Delta y = \left(\frac{\beta_1}{100}\right)\% \Delta x$ One percent change in *x* leads to $\left(\frac{\beta_1}{100}\right)$ unit change in *y*
$\log(y)$	x	**For small change** $\%\Delta y = (100\,\beta_1)\,\Delta x$ One unit change in *x* leads to $(100\,\beta_1)\%$ change in *y* **For large and exact change** $\%\Delta y = 100[\exp(\beta_1 \Delta x) - 1]$

(full version available on paid notes)

Week 1

(1) Types of Data

- Time series: sets of observations taken at different points in time
- Panel data: same individual data is collected over time
- Pooled cross sections: two or more cross sections are combined in one data set

(2) Causal Effect

- Holding all other relevant factors constant, how does variable *y* change if variable *x* changes?
- In practice, it is unfeasible to literally hold 'all else equal', and thus causality can be difficult to establish

(3) Population Regression Function (PRF)

$E(y|x) = E(\beta_o + \beta_1 x + u|x)$
$E(y|x) = \beta_o + \beta_1 x + E(u|x)$
Because of ZCM assumption, $E(u|x) = 0$
$E(y|x) = \beta_0 + \beta_1 x$

- This is the best model we can have, but no models are perfect an hence the population regression model takes another form

(4) Population Regression Model

$$y = \underbrace{\underbrace{\beta_0 + \beta_1 x}_{E(y|x)}}_{systematic} + \underbrace{u}_{Non-systematic}$$

(5) Zero Conditional Mean (ZCM)

- $\mathrm{E(u|x) = E(u)}$ → *u* is mean independent of *x*
  and, $\mathrm{E(u) = 0}$ → The average value of *u* is zero
  $\therefore \mathrm{E(u|x) = 0}$

- This assumption is not very restrictive – the intercept could always be redefined to hold this assumption:

Suppose $E(u) = \alpha_0 = 0$, the model can always be rewritten with the same slope, but a new intercept and error; where the new error has a zero expected value:

$y = \beta_0 + \beta_1 x + u + \alpha_0 - \alpha_0$
$y = (\beta_0 + \alpha_0) + \beta_1 x + (u - \alpha_0)$

$E(u - \alpha_0) = E(u) - E(\alpha_0)$
$E(u - \alpha_0) = \alpha_0 - \alpha_0$
$E(u - \alpha_0) = 0$

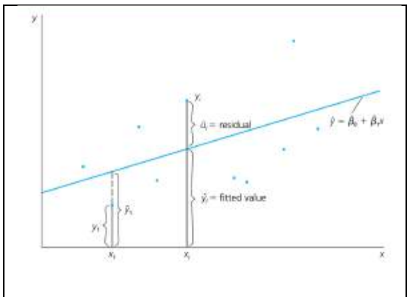
(6) Sample Regression Function

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

(7) Sample Regression Model

$$y = \underbrace{\hat{\beta}_0 + \hat{\beta}_1 x}_{\hat{y}} + \hat{u}$$

- $\hat{u}$ is the estimator of u (residual term)



Week 2

(1) Ordinary Least Squares

- The residual for observation i is the difference between the actual y_i and its fitted:

$$\hat{u}_i = y_i - \hat{y}_i$$
$$\hat{u}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_i x_i)$$

- OLS solves the minimisation problem: $min \sum_{i=1}^{n} \hat{u}_i^2$, to estimate $\hat{\beta}_0$ and $\hat{\beta}_1$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

(2) Algebraic Properties of OLS

- Sample average of residuals is zero, $\sum_{i=1}^{n} \hat{u}_i = 0$
- Covariance between regressors and residuals is zero, $\sum_{i=1}^{n} x_i \hat{u}_i = 0$
- Sample averages of dependent and independent variables lie on the fitted regression line, $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$

(3) Coefficient of determination

- It is the total variation in the dependent variable (y) that is explained by the regression (goodness of fit)

$$SST = SSE + SSR$$

$$\underbrace{\sum_{i=1}^{n}(y_i - \bar{y})^2}_{Total\ Sum\ of\ Squares} = \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}_{Explained\ Sum\ of\ Squares} + \underbrace{\sum_{i=1}^{n} \hat{u}_i}_{Sum\ of\ Squares\ Residuals}$$

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \qquad 0 \le R^2 \le 1$$

- R^2 is equal to the squared correlation between y_i and $\hat{y}_i$
- Very high R^2 is common with time series data
- Low R^2 is common with cross sectional data

(4) 'Linear' models

- Linear regression means a regression that is linear in parameters – the βs are linear

Model	Dependent	Independent	Interpretation of β_1
Level-level	y	x	$\Delta y = \beta_1 \Delta x$ One unit change in *x* leads to a β_1-unit change in *y*
Level-log	y	$\log(x)$	$\Delta y = \left(\frac{\beta_1}{100}\right)\% \Delta x$ One percent change in *x* leads to $\left(\frac{\beta_1}{100}\right)$ unit change in *y*
Log-level	$\log(y)$	x	$\%\Delta y = (100\, \beta_1)\, \Delta x$ One unit change in *x* leads to $(100\, \beta_1)\%$ change in *y*
Log-log	$\log(y)$	$\log(x)$	$\%\Delta y = \beta_1 \%\, \Delta x$ One percent change in *x* leads to a β_1% change in *y*

(5) Gauss-Markov Assumptions – Simple Linear Model

- SLR.1 – Linear in parameters
- SLR.2 – Random sampling
- SLR.3 – Sample variation in the explanatory variable
  - Sample outcomes on *x* are not all the same value
- SLR.4 – Zero conditional mean $\mathrm{(u|x_1) = 0}$
- SLR.5 – Homoskedasticity $Var(u|x_1) = \sigma^2$

- SLR.1 – 4 → OLS estimators are unbiased $\mathrm{E(\hat{\beta}_j) = \beta_j}$
  - Unbiasedness: *on average* the estimated coefficient they will equal to the true parameter.
- SLR.1 – 5 → OLS estimators are
  - Best (minimum variance and most efficient)
  - Linear (linear – SLR.1)
  - Unbiased (SLR.1 – 4)
  - Estimators

(6) Effects of Changing Units of Measurement

- If y is multiplied by a constant c → OLS intercept and slope estimates are also multiplied by c
- If x is multiplied by a constant c → OLS intercept is unchanged but the slope estimate is divided by *c*
- R^2 is the same when varying the units of measurement
-