Introductory Econometrics

Part A: Preliminaries

Introduction to Econometrics

What is Econometrics?

- Econometrics is the use of statistical methods for estimating economic relationships, testing economic theories, and evaluating and implementing government and business policy.

Examples

- Forecasting macroeconomic variables like interest rates, inflation rates and GDP
- Examine the effects of political campaign expenditures on voting outcomes
- Consider the effect of school spending on student performance.
- How will an increase in fares for public transportation affect the number of travellers who switch to car or bike?
- What is the relationship between advertising and sales?
- How much violent crime will be reduced if an additional million dollars is spent putting more police on the street?

Types of Data

- Different statistical techniques are based on different assumptions about the nature of the data
- It is important to be aware of these assumptions to make sure that the techniques we are using are valid for the data we have
- Cross-Section Data: cover a given period of time, e.g. per capita disposable income during some given time period
- Time-Series Data: are collected over discrete intervals of time, e.g. per capita disposable income can be collected over quarterly intervals between 1961-1980

Overview of Probability

Random Variables

- A random variable can be thought of as an unknown value that may change every time it is inspected (chance mechanism)
- Continuous random variables*:* can take on any value in some interval of values

Probability Distributions

- A probability density function or pdf is a graph, table or equation that summarises information about the relative probabilities of different outcomes occurring, 480
- A cumulative distribution function (cdf) is basically the same thing, except that it represents the probability that the random variable is equal to or less than a particular value, 481

Mean, Median and Mode

- The mean is simply the sum of all outcomes divided by the number of outcomes

- The median measures the outcome that is exactly in the middle of the distribution
- The mode measures where the peak of the distribution occurs

Expected Value

- The covariance and correlation between two random variables measures the degree of linear association between the variables, 491-492
- If random variables X and Y are independent, then their covariance and correlation will both be zero, 492
- Note that the converse is not necessarily true, 492

Variance and Standard Deviation

- The variance of a random variable often denoted by σ^2 is a measure of dispersion of the distribution about its mean
- It is defined as: $\sigma^2 = E(X^2) - \mu^2$
- Note that the variance of any constant is zero and if a random variable has zero variance then it is essentially constant

Skewness and Kurtosis

- Skewness measures the lack of symmetry of a distribution. If the distribution is symmetric then skewness = 0
- Kurtosis measures the *peakedness* of a distribution. A distribution with large kurtosis has more values concentrated near the mean and a relatively high central peak
- Usually the kurtosis is reported relative to the kurtosis of the normal distribution, which is three

Covariance

- The covariance between two random variables X and Y is called the population covariance to emphasize that it concerns the relationship between two variables describing a population
- It is a measure of the strength and direction of any linear relationship between X and Y
- If X and Y are independent then Cov(X,Y) = 0, although the converse is not true

Correlation Coefficient

- Correlation has the advantage that its value always lies in the range −1 to +1
- The correlation coefficient between X and Y is given by: $\mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$
- -1 indicates perfect negative association, while 1 indicates perfect positive association

Formulae

The Summation Operator

$$\sum_{i=1}^{n} f(x_i) = f(x_1) + f(x_2) + \cdots + f(x_n) = \sum_{x} f(x)$$

$$\sum ax_i = a \sum x_i$$

$$\sum_{i=1}^{n} a = na$$

$$\sum_{n}(x_i + y_i) = \sum_{n} x_i + \sum_{n} y_i$$

$$\sum_{i=1}(x_i - \bar{\mathrm{x}}) = \sum_{i=1} x_i - \sum_{i=1} x_i = 0$$

Double Summation

$$\sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j = x_1 y_1 + x_1 y_2 + \cdots + x_2 y_1 + x_2 y_2 + \cdots$$

$$= (x_1 + x_2 + \cdots + x_n) y_1 + (x_1 + x_2 + \cdots + x_n) y_2 + \cdots$$

$$= \left(\sum_{i=1}^{n} x_i\right)(\mathrm{y}_1 + \mathrm{y}_2 + \cdots + \mathrm{y}_\mathrm{m})$$

$$= \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{m} y_i\right)$$

Expected Value

$$E(X) = \sum_{x} x f(x)$$

$$E[g(X)] = \sum_{x} g(x) f(x)$$

$$E[g_1(X) + g_2(X)] = E_1[g(X)] + E_2[g(X)]$$

$$E[g(X,Y)] = \sum_{x}\sum_{y} g(x,y) f(x,y) = \sum_{x}\sum_{y} g(x,y) f(x)\, f(y)$$

Variance

$\mathrm{var}(X) = E(X - \mu)^2$

$$= \sum_{x} (x - \mu)^2 f(x)$$

$$= \sum_{x} (x^2 - 2x\mu + \mu^2) f(x)$$

$$= \sum_{x} x^2 f(x) - \sum_{x} 2x\mu f(x) + \sum_{x} \mu^2 f(x)$$

$$= \sum_{x} x^2 f(x) - 2\mu \sum_{x} x f(x) + \mu^2 \sum_{x} f(x)$$

$$= \sum_{x} x^2 f(x) - 2\mu\mu + \mu^2$$

$$= \sum_{x} x^2 f(x) - 2\mu^2 + \mu^2$$

$$= \sum_{x} x^2 f(x) - \mu^2$$

$$\sigma^2 = E(x^2) - E(x)^2$$

For calculation: $\sigma^2 = \frac{1}{n-1}\sum(x-\mu)^2$

Covariance

$$
\begin{aligned}
\operatorname{cov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\
&= E[XY - \mu_Y X - \mu_X Y + \mu_X \mu_Y] \\
&= E[XY] - E[\mu_Y X] - E[\mu_X Y] + E[\mu_X \mu_Y] \\
&= E[XY] - \mu_Y \sum_{i=1}^{n} x_i f(x_i) - \mu_X \sum_{i=1}^{n} y_i f(y_i) + E[\mu_X \mu_Y] \\
&= E[XY] - \mu_Y \mu_X - \mu_X \mu_Y + E[\mu_X \mu_Y] \\
&= E[XY] - \mu_X \mu_Y - \mu_X \mu_Y + \mu_X \mu_Y \\
\sigma_{XY} &= E[XY] - \mu_X \mu_Y
\end{aligned}
$$

Correlation

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Part B: The Basic Regression Model

Introduction to the Regression Model

The Economic Model

- An Economic Model summarizes what theory tells us about the relationship between two variables x and y
- An Economic Model is an abstraction of reality – it will always only be an approximation of the true relationship
- Sources of errors include missing variables due maybe because they include variables that cannot be observed or it is difficult to collect data

Introducing the Error Term

- In order to estimate the true population parameters β_1 and β_2, we must introduce an error term e, which accounts for all sources of error like missing variables, non-linearity, etc
- These factors cause individual observations y to differ from the mean value E(y) $= \beta_1 + \beta_2 x$
- The error term is the random component of the analysis, and represents the difference between the actual observed y and its conditional mean value (for that value of x), 15
- Note that the error term is never actually known, as we never know what the true parameters of the regression are (β_1 and β_2), only their estimates, 17
- Also note that because y and e differ only by a constant ($\beta_1 + \beta_2 x$), they must have the same variance, 15
- The expected value of the error, however, is always zero, 15
- It is also assumed that the covariance between any pair of random errors is zero, 16

Residual versus the Error

- The error is defined as the distance between a particular data point and the true regression line: $e_i = y_i - \beta_1 - \beta_2 x_i$
- Because we don't know $\beta_1$and β_2, the error is not observable
- The residual for observation *i* is the difference between the actual y_i and its fitted value: $\hat{e}_i = y_i - \hat{y}_i = y_i - b_1 - b_2 x_i$

- There are n such residuals and unlike the error they can actually be observed

Estimator Notation

$b_1 = \hat{\beta}_1 = \text{estimator of } \beta_1$

$b_2 = \hat{\beta}_2 = \text{estimator of } \beta_2$

$e_i = y_i - \beta_1 - \beta_2 x_i = \text{error term, difference between actual } y \text{ and true regression line}$

$\hat{e} = y - \hat{y} = y_i - b_1 - b_2 x_i$
$= \text{residual, estimate of the error, difference between } y \text{ and estimated regression line}$

$\hat{y} = y - \hat{e} = \text{fitted value of } y\text{, the estimated value of the regression line}$

Multiple Regression Model

- Although the intercept term of the multiple regression model seldom has any meaningful economic interpretation, we need to include it in the model to make it work, 107
- The coefficients $\beta_2, \beta_3 \ldots \beta_K$ are unknown coefficients corresponding to the explanatory variables $x_2, x_3 \ldots x_K$
- K is the total number of parameters being estimated, including β_1
- A single parameter β_K measures the effect of a change in the variable x_K upon the expected value of y all other variables held constant (i.e. the partial derivative of y relative to x_K)
- The parameters can be thought of representing the slop along different axes of a plane, 108

Six Assumptions of the Model

- 1. $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} \ldots + \beta_K x_{iK} + e$ – model correctly specified
- 2. $E(y_i) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} \ldots + \beta_K x_{iK}$, as $E(e_i) = 0$
- 3. $\text{var}(e_i) = E(e_i^2) = \sigma^2 = \text{var}(y)$ the random variables y and e have the same variance because they differ only by a constant – homoskedasticity
- 4. $\text{cov}(e_i, e_j) = E(e_i e_j) = \text{cov}(y_i, y_j) = 0$ successive errors are uncorrelated and are not clustered together – serial independence of errors
- 5. The variables x_{iK} are not random and are not exact linear functions of the other explanatory variables (i.e. they all impart additional information)
- 6. $e \sim N(0, \sigma^2)$ – errors are normally distributed

The Least Squares Method

Introduction

- In order to estimate the values of the population parameters, we generally use the least squares estimation method, which is arbitrary but very effective, 19-20
- It basically consists of drawing a line (and hence picking estimate values) that minimises the sum of the squared difference of each observation from the line along the y-axis, 20
- The vertical distance between each point and the time are the least squares residuals, which form our estimate of the error term for that value of x, 21
- Note that b_1 and b_2 are referred to as the least squares *estimators*, while the particular values they take are the *estimates*, 22

Deriving the Least Squares Estimator

$$\hat{e} = y - \hat{y}$$
$$= y_i - b_1 - b_2 x_i$$
$$\sum_{i=1}^{n} \hat{e}^2 = \sum_{i=1}^{n} (y_i - b_1 - b_2 x_i)^2$$
$$= \sum_{i=1}^{n} \left(y_i^2 - 2y_i b_1 - 2y_i b_2 x_i + b_1^2 - 2b_1 b_2 x_i + b_2^2 x_i^2\right)$$
$$\frac{\partial(\sum_{i=1}^{n} \hat{e}^2)}{\partial b_1} = \sum_{i=1}^{n} (-2y_i + 2b_1 - 2b_2 x_i) = 0 \qquad (1)$$
$$\frac{\partial(\sum_{i=1}^{n} \hat{e}^2)}{\partial b_2} = \sum_{i=1}^{n} \left(-2y_i x_i + 2b_1 x_i + 2b_2 x_i^2\right) = 0 \qquad (2)$$

From (1)

$$\sum(-2y_i + 2b_1 + 2b_2 x_i) = 0$$
$$\sum(-2y_i) + \sum(2b_1) + \sum(2b_2 x_i) = 0$$
$$-2\sum y_i + 2\sum b_1 + 2b_2 \sum x_i = 0$$
$$-\sum y_i + \sum b_1 + b_2 \sum x_i = 0$$
$$-\sum y_i + nb_1 + b_2 \sum x_i = 0$$
$$nb_1 = \sum y_i - b_2 \sum x_i$$
$$b_1 = \frac{1}{n}\sum y_i - \frac{1}{n} b_2 \sum x_i$$
$$b_1 = \bar{y}_i - b_2 \bar{x}_i \qquad (3)$$

From (2)

$$\sum\left(-2y_i x_i + 2b_1 x_i + 2b_2 x_i^2\right) = 0$$
$$\sum(-2y_i x_i) + \sum(2b_1 x_i) + \sum\left(2b_2 x_i^2\right) = 0$$
$$-2\sum y_i x_i + 2b_1 \sum x_i + 2\sum b_2 x_i^2 = 0$$
$$-\sum y_i x_i + b_1 \sum x_i + \sum b_2 x_i^2 = 0$$

Sub in (3)

$$-\sum y_i x_i + (\bar{y}_i - b_2 \bar{x}_i)\sum x_i + \sum b_2 x_i^2 = 0$$
$$-\sum y_i x_i + (\bar{y}_i - b_2 \bar{x}_i)(n\bar{x}) + b_2 \sum x_i^2 = 0$$
$$\bar{y}_i n\bar{x} - b_2 \bar{x}_i n\bar{x} + b_2 \sum x_i^2 = \sum y_i x_i$$
$$b_2\left(\sum x_i^2 - \bar{x}_i \sum x_i\right) = \sum y_i x_i - n\bar{y}\bar{x}$$
$$b_2\left(\sum x_i^2 - n\bar{x}_i^2\right) = \sum y_i x_i - n\bar{y}\bar{x}$$
$$b_2\left(\sum x_i^2 - n\bar{x}_i^2\right) = \sum y_i x_i - n\bar{y}\bar{x}$$
$$b_2\left(\sum x_i^2 - n\bar{x}_i^2 + n\bar{x}_i^2 - n\bar{x}_i^2\right) = \sum y_i x_i - n\bar{y}_i\bar{x} - (n\bar{y}\bar{x} - n\bar{y}\bar{x})$$
$$b_2\left(\sum x_i^2 - 2n\bar{x}_i^2 + n\bar{x}_i^2\right) = \sum y_i x_i - \sum y_i \bar{x} - \sum x_i \bar{y} + \sum \bar{y}\bar{x}$$
$$b_2\left(\sum x_i^2 - 2n\bar{x}_i \bar{x}_i + n\bar{x}_i^2\right) = \sum(y_i x_i - y_i \bar{x} - x_i \bar{y} + \bar{y}\bar{x})$$
$$b_2\left(\sum x_i^2 - 2\bar{x}_i \sum x_i + n\bar{x}_i^2\right) = \sum(y_i x_i - y_i \bar{x} - x_i \bar{y} + \bar{y}\bar{x})$$
$$b_2 \sum\left(x_i^2 - 2\bar{x}_i x_i + \bar{x}_i^2\right) = \sum(y_i x_i - y_i \bar{x} - x_i \bar{y} + \bar{y}\bar{x})$$
$$b_2 \sum(x_i - \bar{x})^2 = \sum(x_i - \bar{x})(y_i - \bar{y})$$

$$b_2 = \frac{\sum[(x_i - \bar{x})(y_i - \bar{y})]}{\sum[(x_i - \bar{x})^2]}$$

Proving Unbiasedness

- Because b_1 and b_2 are random variables, we can never know how close they are to the actual population parameters, 26
- Rather, all we can do is characterise the estimators in terms of their properties such as mean, variance, covariance, etc, 26
- From the unbiasedness proof shown below, we can determine that the expected value of b_2 is not a random variable, and hence will be an unbiased estimator of the population parameter (note: the assumptions must hold for this to be true), 27-28
- The fact that b_2 is an unbiased estimator of β_2 only means that on average it will produce the correct value – we still don't know how close any particular estimated value of b_2 will be

$$\begin{aligned}
b_2 &= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \\
&= \frac{\sum(x_i - \bar{x})y_i - \sum(x_i - \bar{x})\bar{y}}{\sum(x_i - \bar{x})^2} \\
&= \frac{\sum(x_i - \bar{x})y_i - \sum x_i\bar{y} - \sum\bar{x}\bar{y}}{\sum(x_i - \bar{x})^2} \\
&= \frac{\sum(x_i - \bar{x})y_i - n\bar{x}\bar{y} + n\bar{x}\bar{y}}{\sum(x_i - \bar{x})^2} \\
&= \frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2} \\
&= \sum_{i=1}^{N}\left(\frac{(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} \times y_i\right) \\
&= \sum_{i=1}^{N} w_i y_i \\
&= \sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{\sum(x_i - \bar{x})^2}\right)(\beta_1 + \beta_2 x_i + e_i)
\end{aligned}$$

$$\begin{aligned}
b_2 &= \beta_1\sum w_i + \beta_2\sum w_i x_i + \sum w_i e_i \\
&= \beta_1\sum\left(\frac{x_i - \bar{x}}{\sum(x_i - \bar{x})^2}\right) + \beta_2\sum\left(\frac{x_i - \bar{x}}{\sum(x_i - \bar{x})^2}\right)x_i + \sum w_i e_i \\
&= \beta_1\frac{\sum(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} + \beta_2\frac{\sum(x_i - \bar{x})x_i}{\sum(x_i - \bar{x})^2} + \sum w_i e_i \\
&= \beta_1\frac{\sum x_i - \sum\bar{x}}{\sum(x_i - \bar{x})^2} + \beta_2\frac{\sum x_i^2 - \sum\bar{x}x_i}{\sum x_i^2 - \sum 2\bar{x}x_i + \sum\bar{x}^2} + \sum w_i e_i \\
&= \beta_1\frac{n\bar{x} - n\bar{x}}{\sum(x_i - \bar{x})^2} + \beta_2\frac{\sum x_i^2 - \bar{x}\sum x_i}{\sum x_i^2 - \bar{x}\sum 2x_i + n\bar{x}^2} + \sum w_i e_i \\
&= \beta_1\frac{0}{\sum(x_i - \bar{x})^2} + \beta_2\frac{\sum x_i^2 - \bar{x}n\bar{x}}{\sum x_i^2 - 2\bar{x}n\bar{x} + n\bar{x}^2} + \sum w_i e_i \\
&= \beta_1 \times 0 + \beta_2\frac{\sum x_i^2 - n\bar{x}^2}{\sum x_i^2 - n\bar{x}^2} + \sum w_i e_i
\end{aligned}$$

$$b_2 = \beta_2 + \sum w_i e_i$$

$$\begin{aligned} E(b_2) &= E(\beta_2 + \sum w_i e_i) \\ &= \beta_2 + \sum w_i E(e_i) \\ &= \beta_2 \end{aligned}$$

Using a similar method (see lectures), we can also show that the expected value of b_1 is equal to β_1.

The Gauss Markov Theorem

- The smaller the variance of an estimator is, the greater the sampling precision of that estimator. One estimator is more precise than another estimator if its sampling variance is less than that of the other estimator.
- The least squares principle is only one way of using the data to obtain estimates of b_1 and b_2 - so what do we use it?
- The reason for this is because of the Gauss Markov Theorem, which states that under assumptions 1-5, the estimators $b1$ and $b2$ have the smallest variance of all linear and unbiased estimators of b_1 and b_2
- Note that the theorem does not require that the data be normally distributed, 32
- They are BLUE: Best Linear Unbiased Estimators

Deriving Parameter Variance

- These indicators measure how much b_1 and b_2 are likely to deviate (be spread out around from) their true population parameters, 29
- Deriving equations for these two variances, we can see that they depend upon four variables:
- Population variance: the larger the variance of the underlying population parameter, the larger the variance of its estimators, 30
- Sum of Squares: the larger the sum of the squares of error of the x-values, the lower is the variance of the estimators. This occurs because a higher sum of squares indicates more spread out data (relative to the mean), and hence a broader range of values to make estimates with, 30
- Sample size: the larger the sample size, the smaller the variance of the estimators, 30
- Absolute data size: the farther away the x-values extend from the y-axis, the more will uncertainty in the slope affect the estimation of β_1, and hence the larger the variance of b_1

$$\begin{aligned} \text{var}(X) &= E(X-\mu)^2 \\ \text{var}(b_2) &= E(b_2-\beta_2)^2 \\ \text{var}(b_2) &= E(\beta_2 + \sum w_i e_i - \beta_2)^2 \\ &= E(\sum w_i e_i)^2 \\ &= E(\sum w_i e_i)^2 \\ &= E\left(\sum_i w_i^2 e_i^2 + 2\sum\sum w_i w_j e_i e_i\right) \\ &= \sum w_i^2 E\left(e_i^2\right) + 2\sum\sum w_i w_j E(e_i e_i) \\ \text{var}(b_2) &= \sum w_i^2 E\left(e_i^2\right) + 2\sum\sum w_i w_j E(e_i e_i) \end{aligned}$$

$$NOTE: \sigma^2 = E(x^2) - E(X)^2; \;\; \sigma_{XY} = E[XY] - \mu_X \mu_Y$$

$$\begin{aligned}
\mathrm{var}(b_2) &= \sum w_i^2(\sigma^2 + E(e_i)^2) + 2\sum\sum w_i w_j(\sigma_{w_i w_j} + \mu_{w_i}\mu_{w_j}) \\
&= \sum w_i^2(\sigma^2 + 0) + 2\sum\sum w_i w_j(0 + 0) \\
&= \sigma^2 \sum w_i^2 \\
&= \sigma^2 \sum \left(\frac{x_i - \bar{x}}{\sum(x_i - \bar{x})^2}\right)^2 \\
&= \sigma^2 \sum \frac{(x_i - \bar{x})^2}{(\sum(x_i - \bar{x})^2)^2} \\
&= \frac{\sigma^2 \sum(x_i - \bar{x})^2}{(\sum(x_i - \bar{x})^2)^2} \\
\mathrm{var}(b_2) &= \frac{\sigma^2}{\sum(x_i - \bar{x})^2}
\end{aligned}$$

Unfortunately, the variance of the error term σ^2 is unknown, so instead we use the variance of the residual $\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{N-k}$, where k is the number of variables and N is the number of observations.

$$\widehat{\mathrm{var}}(b_2) = \frac{\hat{\sigma}^2}{\sum(x_i - \bar{x})^2}$$

Using a similar method, we can show that the variance of b_1 is $\mathrm{var}(b_1) = \frac{\sum x_i^2}{N} \times \mathrm{var}(b_2)$.

Variance of the Residual

$$\begin{aligned}
\mathrm{var}(e_i) &= \sigma^2 \\
&= E\left(e_i - \mu_{e_i}\right)^2 \\
&= E(e_i - 0)^2 \\
&= E(e_i)^2 \\
\hat{\sigma}^2 &= E(\hat{e}_i)^2 \\
&= \frac{\sum \hat{e}_i^{\,2}}{N-2}
\end{aligned}$$

Scaling the Data

What is Data Scaling?

- Changes in the units of measurement of variables is called scaling the data
- The choice of the scale is made by the researcher to make interpretation meaningful and convenient
- However, it is crucial to note that the choice of the scale will affect the interpretation of the coefficients

Changing the Scale of X

Suppose the units of measurement of *x* is changed by dividing it by a constant *c*. To compensate for this, it is necessary for us to multiply β_2 by the same constant c.

$$\begin{aligned}
y &= \beta_1 + \beta_2 x + e \\
&= \beta_1 + (c\beta_2)\left(\frac{x}{c}\right) + e
\end{aligned}$$

$$= \beta_1 + \beta_2^* \left(\frac{x}{c}\right) + e$$

Hence if we scale x by a factor of $\frac{1}{c}$, the new coefficient β_2^* will equal $c\beta_2$.

$$se(\beta_2^*) = c \times se(\beta_2)$$

Changing the Scale of Y

In this case we have to divide or multiply all factors of the equation for y by the same constant c, as well as the standard errors of both terms. Thus, changes in the y scale effect both axes.

$$y = \beta_1 + \beta_2 x + e$$
$$\frac{y}{c} = \frac{\beta_1}{c} + \frac{\beta_2}{c} x + \frac{e}{c}$$
$$= \beta_1 + \beta_2^* \left(\frac{x}{c}\right) + e$$

Hence if we scale y by a factor of $\frac{1}{c}$, the new coefficients β_k^* will all be will equal $\frac{\beta_k}{c}$.

$$se(\beta_k^*) = \frac{1}{c} \times se(\beta_k)$$

Note that changes in functional form (e.g. $\frac{y}{c} = \frac{\beta_1}{c} + \frac{\beta_2}{c}\log(x) + \frac{e}{c}$), do not make any difference to this; scaling is done in the same way.

Functional Form

Choosing a Functional Form

- In the real world, many relationships between variables are not linear, 86
- However, by transforming the variables within our linear model, we are able to still use this same model to analyse such non-linear relationships, 86
- Log-log models are particularly convenient, as the elasticity will be constant, 87
- Given the large number of possible functional forms that we could use, many of which have similar shapes, we should try to chose the form that both fits economic theory and the data as closely as possible, whilst ensuring that assumptions SR1-6 are satisfied, 88

Elasticity and Interpretation

The elasticity ϵ_{yx} refers to the percentage change in y associated with a 1% change in x. Interpretation is not necessarily the same as elasticity. Interpretation is simply a meaningful and simple way to describe the change in one variable y relative to another x. This will take a different form for different functional forms, and could be elasticity (as in log-log), simple absolute changes (linear-linear), or a hybrid form (see log-linear and linear-log models).

Quadratic Model

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + e_i$$

$$\varepsilon_{yx} = \frac{dy}{dx} \times \frac{x}{y}$$

$$= (\beta_2 + 2\beta_3 x) \times \frac{x}{y}$$

$$\text{elasticity} = (\beta_2 + 2\beta_3 x)\frac{x}{y}$$

- One main reason for using a quadratic specification is to capture turning points
- Turning points occur when the effect of an additional unit of x_i causes a change in the direction of the effect of x_i on y_i
- To find the turning point of the parabola, we simply differentiate $\hat{y}$ with respect to x and equate to zero

The Reciprocal

$$y = \beta_1 + \beta_2(1/x) + e$$

$$\varepsilon_{yx} = \frac{dy}{dx} \times \frac{x}{y}$$

$$= -\beta_2\left(\frac{1}{x^2}\right) \times \frac{x}{y}$$

$$= -\beta_2 \frac{1}{xy}$$

$$\text{elasticity} = \frac{-\beta_2}{xy}$$

- As *x* increases indefinitely the term $\beta_2\left(\frac{1}{x}\right)$ approaches 0 and *y* approaches the limiting or asymptotic value β_1
- Therefore the reciprocal model has an asymptote or limit value that the dependent variable will take when the value of *x* increases indefinitely

Log-Log Function

$$\log(y) = \log(\beta_1) + \beta_2 \log(x) + e$$

$$y = e^{\log(\beta_1)+\beta_2\log(x)+e}$$

$$\varepsilon_{yx} = \frac{dy}{dx} \times \frac{x}{y}$$

$$= \frac{\beta_2}{x} e^{\log(\beta_1)+\beta_2\log(x)+e} \times \frac{x}{y}$$

$$= \frac{\beta_2}{x} y \times \frac{x}{y}$$

$$= \frac{\beta_2 y}{x}\frac{x}{y}$$

$$\text{elasticity} = \beta_2$$

- For log-log models to work, both betas must be positive

Log-Linear Function

$$\log(y) = \beta_1 + \beta_2 x + e$$

$$y = e^{\beta_1+\beta_2 x+e}$$

$$\varepsilon_{yx} = \frac{dy}{dx} \times \frac{x}{y}$$
$$= \beta_2 e^{\beta_1 + \beta_2 x + e} \times \frac{x}{y}$$
$$= \beta_2 y \times \frac{x}{y}$$
$$\text{elasticity} = x\beta_2$$

Interpretation:

$$\Delta \log(y) = \beta_2 \Delta x$$
$$100(\log(y_1) - \log(y_0)) = 100\beta_2 \Delta x$$
$$100\left(\frac{\Delta y}{y_1}\right) \cong 100\beta_2 \Delta x$$
$$\%\Delta y \cong 100\beta_2 \Delta x$$

Linear-Log Function

$$y = \beta_1 + \beta_2 \log x + e$$

$$\varepsilon_{yx} = \frac{dy}{dx} \times \frac{x}{y}$$
$$= \frac{\beta_2}{x} \times \frac{x}{y}$$
$$\text{elasticity} = \frac{\beta_2}{y}$$

Interpretation:

$$\Delta y = \beta_2 \Delta \log(x)$$
$$100\Delta y = 100\beta_2(\log(x_1) - \log(x_2))$$
$$100\Delta y \cong 100\beta_2 \frac{\Delta x}{x_1}$$
$$100\Delta y \cong \beta_2 100 \frac{\Delta x}{x_1}$$
$$\Delta y \cong \frac{\beta_2}{100} \times \%\Delta x$$

Mnemonic: whichever of x or y has the log in front of it becomes interpreted as a percentage change, while the one without the log is multiplied by 100.

Part C: Extensions and Applications of the Model

Confidence Intervals

Purpose of Intervals

- The purpose of interval estimation is to develop a range of values within which the parameters b_1 and b_2 are likely to fall, 49

- Interval estimates are a convenient way to present regression results because they combine point estimation with a measure of the sampling variability, 53-54
- These intervals are referred to as confidence intervals or interval estimates, 49

Finding Intervals

- If we make the assumption that errors are normally distributed, then the probability distributions of the estimators will also be normally distributed, 32-33
- Also, even if the errors are not normally distributed, the Central Limit Theorem tells us that the least squares estimators approximate a normal distribution for a sufficiently large sample size, 33

Introducing the Standard Normal

- Using this assumption of normality, we derive the following estimator for β_2:

$$b_2 \sim N\left(\beta_2, \frac{\sigma^2}{\sum(x_i - \bar{x})^2}\right)$$

- To make this easier to work with, we can define a new variable Z such that:

$$Z = \frac{b_2 - \beta_2}{\sqrt{\sigma^2 / \sum(x_i - \bar{x})^2}} \sim N(0,1)$$

Confidence Interval for b_2

Using this new standard normal distribution, we can derive a confidence internal for b_2:

$$P(-1.96 \leq Z \leq 1.96) = 0.95$$

$$P\left(-1.96 \leq \frac{b_2 - \beta_2}{\sqrt{\sigma^2 / \sum(x_i - \bar{x})^2}} \leq 1.96\right) = 0.95$$

$$P\left(-1.96\sqrt{\sigma^2 / \sum(x_i - \bar{x})^2} \leq b_2 - \beta_2 \leq 1.96\sqrt{\sigma^2 / \sum(x_i - \bar{x})^2}\right) = 0.95$$

$$P\left(-1.96\sqrt{\sigma^2 / \sum(x_i - \bar{x})^2} - b_2 \leq -\beta_2 \leq 1.96\sqrt{\sigma^2 / \sum(x_i - \bar{x})^2} - b_2\right) = 0.95$$

$$P\left(1.96\sqrt{\sigma^2 / \sum(x_i - \bar{x})^2} + b_2 \geq \beta_2 \geq -1.96\sqrt{\sigma^2 / \sum(x_i - \bar{x})^2} + b_2\right) = 0.95$$

$$P\left(b_2 - 1.96\sqrt{\sigma^2 / \sum(x_i - \bar{x})^2} \leq \beta_2 \leq 1.96\sqrt{\sigma^2 / \sum(x_i - \bar{x})^2} + b_2\right) = 0.95$$

The T-Distribution

- However, this is not the whole story, because we don't actually know what σ^2 is, it being an unknown population parameter
- Hence we need to replace it with its estimate, $\hat{\sigma}^2 = \frac{\sum e_i^2}{\mathrm{N}-2}$
- Doing so alters the variance of the distribution, and hence requires us to replace the original standard normal distribution with a new student t-distribution with degrees of freedom N-2
- Note that when N is very large, the t distribution reduces down to the normal distribution
- Otherwise, it is just like a normal distribution but with a larger variance, reflecting our uncertain knowledge of the real value of σ^2; this is why it is used, as it reflects the additional uncertainty created by using a variance estimate

$$t = \frac{b_2 - \beta_2}{\sqrt{\hat{\sigma}^2 / \sum(x_i - \bar{x})^2}} = \frac{b_2 - \beta_2}{\sqrt{\widehat{\mathrm{var}}(b_2)}} = \frac{b_2 - \beta_2}{\widehat{\mathrm{se}}(b_2)} \sim \mathrm{t}_{\mathrm{N}-2}$$

The Meaning of Confidence Intervals

- If we construct a standard 95% confidence interval, it means that 95% of intervals constructed through repeated sampling will contain the true value of b_2, 50
- The size of the confidence interval depends upon what critical value we select – the larger the t_c, the more likely it is to contain the actual value, but the less specific is the interval

Confidence Interval Formula

Confidence interval for b_j given by: $b_j \pm t_c \text{se}(b_j)$

Hypothesis Testing

The Basic Idea

- Conducting a hypothesis tests requires making an assumption that the null hypothesis is true, and that hence the probability distribution is a specific value, 55
- To determine the likelihood that the null is true, we construct a confidence interval on the assumption that it is, 55-56
- The idea is that if the value of the test statistic falls in a region of low probability, then it is unlikely that the test statistic has the assumed distribution, and hence it is unlikely that the null is true, 56

The Null Hypothesis

- A null hypothesis is the belief we will maintain until we are convinced by the sample evidence that it is not true, in which case the null hypothesis is rejected
- $H_0: \beta_k = c$, where k=1 or 2 and *c* is a constant
- Often we will use c=0, which asks if the coefficient should even be in the model or not

The Alternative Hypothesis

- A logical alternative to the null hypothesis that can be either 1-sided or 2-sided
- If theory suggests that a population coefficient should be a particular sign then we would use a 1-sided test
- One-sided Test – $H_1: \beta_k > c$ or $H_1: \beta_k < c$
- Two-sided Test – $H_1: \beta_k \neq c$
- Rejecting the null hypothesis that $\beta_k = c$ leads us to accept the conclusion that β_k takes a value either larger or smaller than *c*

Types of Errors

- Rejecting a true null is called a Type I error, and its probability is given by the level of significance we chose to use, 56
- A type II error occurs when we fail to reject a false null, and its probability is unknowable, because by definition we do not know what the real distribution looks like, 56

Rejection Regions

- If the alternative hypothesis is that $b_k > c$, then the value of the test statistic will tend to be larger if the alternative hypothesis is true, 56
- Thus, if the t statistic is higher than a specified 'critical value', we conclude that it is highly unlikely that the null is true, so we reject it, 56

- Similarly, if the alternative hypothesis is that $b_k < c$, then the value of the test statistic will tend to be smaller if the alternative hypothesis is true, 57
- Thus, if the t statistic is smaller than a specified 'critical value', we conclude that it is highly unlikely that the null is true, so we reject it, 56
- The null hypothesis is rejected if the t statistic is larger (in absolute value terms, depending on if it's a one or two tailed test) than the critical value

Test Procedure

$H_0: \beta_2 \geq 0$
$H_A: \beta_2 < 0$
Test statistic: $t = \frac{b_1}{\text{se}(b_1)} \sim t_{(N-k)}$
Critical value: $t_c = t_{(1-\alpha, N-K)}$
Decision rule: reject H_0 if $t \leq t_c$
Make decision: $t = x < t_c = y \therefore$ reject or do not reject H_0
Conclusion: "We reject H_0 at the 5% level of significance (or to a 95% level of confidence) since $t > t_c$ or equivalently the p-value < 0.05"

P-Values

- The p-value may be defined as the lowest significance level at which a null hypothesis can be rejected
- As a rule, the smaller the *p* value the stronger evidence against the null hypothesis
- Note that p-values from EViews are reported for a 2-sided alternative hypothesis
- To compute the appropriate p value for a 1-sided alternative, divide the corresponding p value for the 2-sided alternative in half (as half the probability is in each tail).

Forecasting and Prediction

What is Prediction?

- Prediction is the process of using our estimated regression model to predict additional unknown values of y for given values of x, 76
- The prediction interval is a range of values in which the unknown value of y is likely to be located, 76
- As the expected value of the error term is zero, the predictor for the value y_0 will be given by $\hat{y}_0 = b_1 + b_2 x_0$, 77

The Forecast Error

- The forecast error is simply the difference between our actual y value and our predicted y, and is given by: $f = y_0 - \hat{y}_0$, 77
- It turns out that y_0 is the best linear unbiased predictor (BLUP) for y_0, 77

Proving Unbiasedness

Note that we can show that $\hat{y}_0$ is an unbiased estimator of y_0.

$$\begin{aligned} E(f) &= E(y_0 - \hat{y}_0) \\ &= E\big((\beta_1 + \beta_2 x_0) - (b_1 + b_2 x_0)\big) \\ &= E\big((\beta_1 + x_0 + e_0) - (b_1 + b_2 x_0)\big) \end{aligned}$$

$$
\begin{aligned}
&= E(\beta_1 + \beta_2 x_0 + e_0) - E(b_1 + b_2 x_0) \\
&= E(\beta_1 + \beta_2 x_0 + e_0) - E(b_1 + b_2 x_0) \\
&= E(\beta_1 + \beta_2 x_0) + \mathrm{E}(e_0) - E(b_1 + b_2 x_0) \\
&= E(\beta_1 + \beta_2 x_0) + 0 - E(b_1) - E(b_2)x_0 \\
&= \beta_1 + \beta_2 x_0 - \beta_1 - \beta_2 x_0 \\
E(f) &= 0
\end{aligned}
$$

Forecast Error Variance

The variance of the forecast error is given by:

$$
\begin{aligned}
\mathrm{var}(f) &= \sigma^2 \left[1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right] \\
\widehat{\mathrm{var}}(f) &= \hat{\sigma}^2 \left[1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right] \\
&= \left[\hat{\sigma}^2 + \frac{\hat{\sigma}^2}{N} + \frac{\hat{\sigma}^2 (x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right] \\
&= \left[\hat{\sigma}^2 + \frac{\hat{\sigma}^2}{N} + (x_0 - \bar{x})^2 \times \widehat{\mathrm{var}}(b_2)\right]
\end{aligned}
$$

The variance of the forecast error is smaller when:

- The overall variance of the model σ^2 is lower
- The sample size N is larger
- The predicted value is closer to the mean, as measured by $(x_0 - \bar{x})^2$
- The overall sample size of the explanatory variable is more spread out

Prediction Interval

- The predicted value does not tell us how accurate this prediction procedure is likely to be
- This can be done using the prediction interval of y_0 can be found by $\hat{y}_0 \pm t_c \mathrm{se}(f)$, 78
- The prediction interval is essentially the confidence interval for our β_2 parameter adjusted to reflect the fact that values further from the mean (of x and y) are less certain than closer values (i.e. confidence intervals get wider away from the mean)
- If the variance of the forecast error is similar to the estimated variance of the error term, it indicates that the primary uncertainty is coming from the model itself, and hence it may be necessary to include more explanatory variables, 79

Goodness of Fit

Introduction

It is useful to develop a measure of how much of the variation in y_i is explained by the model. We do this by subtracting y from both sides to decompose $y_i - \bar{y}$ into a part that is "explained" by the regression model $(\hat{y}_i - \bar{y})$ and a part that is unexplained $\hat{e}_i$.

$$
\begin{aligned}
y_i - \hat{y} &= (\hat{y}_i - \bar{y}) + \hat{e}_i \\
\sum(y_i - \bar{y})^2 &= \sum(\hat{y}_i - \bar{y})^2 + \sum \hat{e}_i^2 \\
SST &= SSR + SSE
\end{aligned}
$$

Total Sum of Squares

This is a measure of total variation in *y* about the sample mean. SST = 'sum of squares total'.

$$SST = \sum(y_i - \bar{y})^2$$

Explained Sum of Squares

This is the part of total variation in *y* about the sample mean that is explained by or due to the regression. SSR = 'sum of squares due to regression'.

$$SSR = \sum(\hat{y}_i - \bar{y})^2$$

Sum of Squared Errors

That part of total variation in y about its mean that is not explained by the regression. SSE = 'sum of squares of error/residuals'.

$$SSE = \sum \hat{e}_i^2$$

Coefficient of Determination

The coefficient of determination R^2 is the proportion of variation in *y* explained by *x* within the regression model. It is always between zero and one.

$$R^2 = \frac{SSR}{SST} = 1 - \left(\frac{SSE}{SST}\right)$$

The Adjusted R^2

It can be shown that R^2 will never decrease when another variable is added to a regression equation. One solution is to use $\bar{R}^2$ which is an R^2 adjusted for degrees of freedom.

$$\bar{R}^2 = 1 - \frac{N-1}{N-K}(1 - R^2)$$
$$= 1 - \frac{SSE/N-K}{SST/(N-1)}$$

Whether $\bar{R}^2$ rises or falls depends on whether the contribution of the new variable to the fit of the regression more than offsets the correction for the additional degrees of freedom.

Significance of Regression and Linear Restrictions

Testing the Significance of the Regression

- A hypothesis that is often used in multiple regression models involves testing the significance of the regression equation as a whole
- This is a joint test of the hypothesis that all the coefficients except the constant term are zero; this is a joint test and is not equal to a series of independent t tests
- Testing this hypothesis requires use of the F-statistic

The F-Test

- A single null hypothesis is one which has a single restriction on one or more parameters, 136
- A joint null hypothesis is one with two or more restrictions on two or more parameters, 136
- To test a joint null hypothesis, we must use an F test, 136

- Note that F tests can only be used for two-tailed tests, where the alternative hypothesis is in the form ≠ x
- When only a single equality relation with the null hypothesis, the t and F tests are equivalent, while the F test is the only one that will work for testing multiple relations at once, 138

The Concept Behind the F-Test

- The aim of the test is to determine if there is a significant difference in the fits of the constrained equation (that is, when the restrictions are imposed) and the unconstrained equation
- By comparing the SSE with the SSER and adjusting for the degrees of freedom and the number of elements of the model being restricted (J), we construct what is called an F-statistic, 136
- If the null hypothesis is true, the F-statistic will have an F-distribution, 136
- The larger the difference in SSE between the two models, the greater the value of the F statistic, and hence the less likely it is that the null hypothesis is true
- As such, we conduct an hypothesis test whereby the null is rejected if the F-statistic exceeds a certain critical value, 140

Testing Linear Restrictions

- For some economic models, it is necessary to impose linear restrictions on the parameters
- An example is estimating the Cobb-Douglas production function assumption of constant returns to scale
- These restriction can be written in general as βi + βj = M, where M is a constant which would equal unity in the case of constant returns to scale
- The null and alternative hypotheses are respectively $H_0: \beta_i + \beta_j = M$ and $H_1: \beta_i + \beta_j \neq M$
- This hypothesis is tested using an F-statistic: $F = \frac{(SSE_R - SSE_U/q)}{SSE_U/(N-K)}$, where q is the number of restrictions in the restricted model (in this case one)
- The aim of the test is to determine if there is a significant difference in the fits of the constrained equation (when the restrictions are imposed) and the unconstrained equation
- In this sense, it works just like a whole-model F-test, in that the higher the F-statistic, the less likely it is that the null is true

Example Hypothesis Test

$H_0: \beta_2 = \beta_3 = \beta_4 = 0$
H_A: At least one of β_2, β_3 or β_4 $\neq 0$
Test statistic: $F = \frac{R^2/(K-1)}{(1-R^2)/(N-K)} = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} \sim F_{(J,N-K)}$
Critical value: $F_c = F_{(1-\alpha,J,N-K)}$
Decision rule: reject H_0 if $|F| \geq F_c$
Make decision: $F = x > F_c = y$ ∴ reject or do not reject H_0
Conclusion: To a 90% degree of confidence we can conclude that xyz.

K = Number of variables including the intercept
N = Sample size
J = Number of linearly independent restrictions (equal to K-1 in case of whole model test)

Part D: Dummy Variables

The Basic Concept

Defining Dummy Variables

- In empirical work we must often incorporate *qualitative* factors into regression models. For example, the gender or race of an individual, industry of a firm (e.g. manufacturing, retail etc) and location (e.g. states of Australia)
- Qualitative factors often come in the form of binary information
- In econometrics, binary variables are usually referred to as dummy variables; a zero or one

A Single Dummy Independent Variable

- Intercept dummy variables affect the intercept of the graph, while slope dummy variables affect the slope, 171-172
- Intercept dummy variables take the parameter δ, while slope dummy variables take the parameter γ, 171-172
- Despite the fact that dummy variables can take only one of two values, the least squares estimation procedure is unaffected by their inclusion – they are treated like any other variable, 171-172
- For example consider the following model of wage determination:

$$\text{wage} = \beta_0 + \delta_0\text{female} + \beta_2\text{educ} + \text{e}$$

- If female = 1 the person is a female and if female = 0 the person is male
- As such, we are essentially estimating two different models, one for male and one for female, with the same slope but different intercepts (to the value of δ)

The Dummy Variable Trap

- Basically this trap occurs when we include as many dummy variables as there are possible separate categories
- None of the independent variables in our model can be an exact linear transformation of any other, as the least squares estimation procedure fails in such cases, 177-178
- As a result, we cannot ever have two or more dummy variables that are always guaranteed to add to one, because this will be an exact linear transformation of the β_1 parameter, for which x always equals one, 178
- The reason such a model cannot be estimated is because there is an exact linear relationship between two of the parameters to be estimated, and hence we cannot distinguish between them – we can only estimate their sum

Proof of Dummy Variable Trap

$$\hat{y}_i = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 M$$

$$X + Z + M = 1$$

$$\begin{aligned}
\therefore \hat{y}_i &= \beta_0 + \beta_1 X + \beta_2 Z + \beta_3(1 - X - Z) \\
&= \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 - X\beta_3 - Z\beta_3 \\
&= \beta_0 + \beta_3 + \beta_1 X - X\beta_3 + \beta_2 Z - Z\beta_3 \\
&= (\beta_0 + \beta_3) + (\beta_1 - \beta_3)X + (\beta_2 - \beta_3)Z
\end{aligned}$$

Avoiding the Dummy Variable Trap

- There are various ways of resolving this problem, but the simplest one is to use only one dummy variable if there are two classes of the qualitative variable
- The category not included in the model is referred to as the base or reference group
- The intercept term marks the value that occurs for the base group
- The other coefficient term δ refers to the difference in the intercept between the base group and that particular dummy variable category
- It does not matter which category we choose as the base group, so long as we are consistent and careful with our interpretation

Extensions

Interpretation with Logs

- The challenge here is knowing how to interpret coefficients on dummy explanatory variables when the dependent variable is log(y)
- What we have to do is calculate the percentage change in the dummy variable coefficient over the reference group
- This can be found by: $\frac{\delta_1 - \delta_2}{\delta_1}$
- Solving this we find that $\%\Delta(\text{over base}) = 100\ (e^{\hat{\delta}} - 1)$, where δ is the coefficient on the dummy variable
- Note that interpretation with dummy variables is always stated as the change in the dependant variable relative to the base condition

Regression with More Than Two Classses

- Using the rule that the number of dummies must be one less than the number of categories of the variable, we can add as many classes to our dummies as we like
- For example, the variable *education* is qualitative in nature being described by three mutually exclusive levels of education: less than high school, high school and university
- Hence we define the model: $y = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta x_i + e$, where x_i is annual income
- Essentially, therefore, each level of education is permitted to have a different intercept term
- After estimation we could test the hypotheses that $\alpha_2 = 0$, $\alpha_3 = 0$ or $\alpha_2 = \alpha_3 = 0$
- Note that it does not matter which dummy variable category is excluded, as we can easily transform the regression from one to another using simple substitution and re-arrangement of the terms
- Such substitution and re-arrangement is based on the simple fact that $D_{1i} + D_{2i} + D_{3i} = 1$
- It is important, however, that one remembers to interpret the estimates in terms of the dummy variable that is excluded

Regression with Several Qualitative Variables

- We can use several dummy independent variables in the same equation, several of which may have more than one category
- Each variable should have one less dummy parameter than the number of categories it includes
- The base group is defined as that when all the dummy variables that are included in the model are equal to zero

- The usual practice to test the significance of dummy variables is to test each category as a group, using an F test if the group has more than two categories

Interactions Between Variables

- Normally, the effects of different parameters in a multiple regression are additive, 175-176
- Sometimes, however, we may suspect that some of these variables may interact with each other, 176
- For example, in estimating a model of wages compared to race and gender, we may suspect that being black and a female would have an additional effect on wages above and beyond the individual effects of either of these operating individually, 176
- To incorporate this into the model, we can include an additional dummy variable consisting of the original two dummies multiplied together, such that its effect is only felt when both dummies are positive, 176
- It is possible when this is done that of the original lone variables will no longer be significant, indicating that its effect is solely indirect, through its interaction with the other variable, 184

Further Applications

Dummy Variables in Seasonal Analysis

- Many economic time series based on monthly or quarterly data exhibit seasonal patterns (regular oscillatory movement)
- Examples are sales of department stores at Christmas time, demand for ice cream and soft drinks during the summer etc
- Often it is desirable to remove the seasonal factor or component from a time series so that one may concentrate on the other components
- The process of removing the seasonal component from a time series is known as deseasonalization or seasonal adjustment
- The time series thus obtained is called the deseasonalized or seasonally adjusted time series

Number of Seasonal Dummies to Include

- The number of seasonal dummy variables to include depends on the nature of the data
- For monthly we needed 11 and a constant
- For quarterly data, 3 dummy and a constant would need to be included
- We can do an F test on the entire model to determine if seasonality is relevant

Testing for Structural Change

- The relationship between the dependent and independent variables may undergo a structural change (also referred to as structural instability or structural breaks), perhaps due to a policy change such as the introduction of the GST or a stock market crash
- One way of testing for structural change is to introduce "dummy variables" and test the significance of their coefficients
- The dummy variable will be defined as ‘turning on’ after we suspect that the structural change has occurred
- We can test the significance of each of these dummies individually, or test them all together (i.e. that all structural change dummies equal zero) in a special F test called a chow test

Part E: Problems with the Model

Assumptions One and Two: Model Specification

What is Model Specification?

- Model specification relates to choosing the *functional form* of the model, and then selecting the *explanatory variables* to include, 148
- Equation specification errors include (1) omission of important variable(s) and (2) inclusion of superfluous variable(s)
- Model specification errors include (1) adopting a wrong functional form or (2) using an incorrect specification of the disturbance term

Omitted Variables

- Omission from the model of a variable that is relevant is known as the emitted-variable bias, as it causes the estimated coefficients of other parameters to become biased (too high or too low)
- It can be viewed as imposing an incorrect constraint on the parameters of the form that $\beta_i = 0$
- It should be noted, however, that the size of the omitted-variable bias depends on how strongly correlated the emitted variables are with existing variables in the model; the closer the correlation, the greater the bias if they are emitted, 150

Irrelevant Variables

- Inclusion of irrelevant variables (those with high p-values) will complicate the model unnecessarily, 150-151
- Also, including irrelevant variables will increase the variance of the estimated values of relevant parameters, thereby reducing the accuracy of the model, 151

Incorrect Functional Form

- The most obvious situation is where the true population regression equation is nonlinear but a linear estimating equation is adopted
- When we use an incorrect functional form, it is likely that we will have some kind of bias in the estimation
- Also, use of an incorrect functional form increases the chances of mistaken inferences about the true population parameters

Choosing The Model

- There are a number of criteria we can use to help decide whether a given variable belongs in the equation:
- 1. Choose variables and functional form on the basis of your theoretical and general understanding of the relationship.
- 2. One method for assessing whether a variable should be included in an equation is to use t-tests. Failure to reject the hypothesis that $\beta_i = 0$can be an indication the variable is irrelevant
- 3. We can use the adjusted R-squared to see if the overall fit of the equation improves when the variable is added to the equation

- 4. Bias: Do other variables' coefficients change significantly when the variable is added to the equation? If they do, this indicates that the results of the regression may not be robust
- However, we must use careful judgement and should not rely on a single criterion to determine the specification. The single most important determinant of a variable's relevance is its theoretical justification.

The RESET Test

- The RESET test is a very useful method of detecting an incorrectly specified model, 152
- The test involves incorporating polynominal (e.g. quadratic or cubic) transformations of the estimated dependent variable (y) values of the initial model as artificial parameters in a new version of the model, 152
- We then test the coefficients of these parameters with a null that they equal zero, 152
- If we can reject this null, this is an indication that the fitted term is adding to the model, and hence we conclude that our original model is likely to be inadequate (mispecified)
- This test can also pick up omitted variables that are correlated with included variables, as if they are correlated with included variables they are likely to be correlated with the artificial variables that have just been introduced, 152

Conducting the RESET Test

The Models:

$$\hat{y}_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + e_i$$
$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \gamma_1 \hat{y}_i^2 + \gamma_2 \, \hat{y_i}^3 + e_i$$

Hypotheses:

$$H_0: \gamma_1 = \gamma_2 = 0$$
$$H_1: \text{at least one of } \gamma_i \neq 0$$

The test-statistic is given by:

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} \sim F_{(J,N-K)}$$

Where $K =$ Number of variables including the intercept, $N =$ Sample size, $J =$ Number of linearly independent restrictions (equal to K-1 in case of whole model test).

Rejection the null indicates the artificial model is valid, which lends support to the idea that the original model was misspecified.

Assumption Three: Heteroskedasticity

What is Heteroskedasticity?

- One of the assumptions of OLS was that the disturbances *e* were identically distributed with mean zero and equal variance
- This assumption is known as homoskedasticity. In this case all the observations can be thought of as being drawn from the same distribution

- Heteroskedasticity refers to the situation whereby the variance of the random variable e (the error term) changes as x changes, 198
- This means that the probability distribution of e and hence of that corresponding y becomes more spread out, 198
- Heteroskedasticy often occurs in data sets in which there is a wide disparity between the largest and smallest observed value of the dependent variable, such as cross-sectional data

Consequences of Heteroskedasticity

- If OLS is used to estimate the coefficients in an instance of Heteroskedasticity, then the following properties hold:
- 1. the OLS estimates is still a linear and unbiased estimator.
- 2. the OLS estimates are no longer BLUE and will be inefficient. That is, there is another estimator with a smaller variance; the 'generalised least squares estimator'
- 3. the standard errors usually computed for the least squares estimator are incorrect. Hence confidence intervals and hypothesis tests that use these standard errors may be misleading
- In order to adjust for this, it is necessary to use a more complicated formula to calculate the 'White standard errors', also called the 'Heteroskedasticity robust' standard errors, 201
- Use of this formula will actually produce narrower confidence intervals than if we merely ignored Heteroskedasticity, 202

Weighted/Generalised Least Squares

- If the error term varies with observations, we will need to find a way of adjusting for this variation such that it 'cancels out', and the adjusted error term is a constant again
- It can be shown that this can be done by dividing each error term by the square root of the function of e in terms of x_i
- Estimators produced using this procedure are called weighed least squares estimators, as they are weighted according to the value of x, 204
- This process of transformation only changes the value of the variance, and does not affect the interpretation of the coefficients (as the same transformation was applied to both sides)
- The GLS estimates satisfy all the standard assumptions and are hence *BLUE*

$$\text{Assume that } \mathrm{var}(e_i) = \sigma^2 h(x_i)$$
$$E(e_i - E(e_i))^2 =$$
$$E(e_i - 0)^2 =$$
$$E(e_i)^2 =$$
$$E(e_i^2) = \sigma^2 h(x_i)$$
$$\frac{1}{h(x_i)} E(e_i^2) = \sigma^2$$
$$E\left(\frac{e_i^2}{h(x_i)}\right) = \sigma^2$$
$$E\left(\left(\frac{e_i}{\sqrt{h(x_i)}}\right)^2\right) = \sigma^2$$

Hence the new equation becomes

$$\frac{y_i}{\sqrt{h(x_i)}} = \frac{\beta_1}{\sqrt{h(x_i)}} + \beta_2\left(\frac{x_{i2}}{\sqrt{h(x_i)}}\right) + \beta_3\left(\frac{x_{i3}}{\sqrt{h(x_i)}}\right) + \frac{e_i}{\sqrt{h(x_i)}}$$

Heteroskedasticity-Consistent Standard Errors

- To apply weighted least squares we need to know the form of heteroskedasticity so we know what weights to apply
- An alternative approach that does not rely on the form of heteroskedasticity being known is to use heteroskedasticity-consistent standard errors (also known as robust standard errors or White standard errors)
- We know that if heteroskedasticity is present the OLS estimates are unbiased, but the corresponding standard errors computed for the least squares estimator are incorrect
- Hence, the idea of the white test is to use the OLS estimates but improve the estimation of the standard errors
- Thus heteroskedasticity-consistent standard errors are standard errors that have been calculated to avoid the consequences of heteroskedasticity, and so are generally more accurate than uncorrected standard errors for large samples in the presence of heteroskedasticity
- They will not, however, be as good as the weighted least squares intervals

Graphing Test for Heteroskedasticity

- One way of investigating the existence of heteroskedasticity is to estimate your model using OLS and to plot the least squares residuals
- If the errors are homoskedastic there should be no patterns of any sort in the residuals
- If errors are heteroskedastic they may tend to exhibit greater variation in some systematic way
- Another method is the goldfeld-quant test, which divides the sample up into two sub-samples, and examines if there is a difference in the variances of their error terms, 211

The White Test for Heteroskedasticity

- The white test is a very useful heteroskedasticity test which involves constructing a model that attempts to explain the error-square terms by changes in the explanatory variables of the original equation, as well as their squares and cross-products, 214-215
- Note that we also include some additional terms, which will be the squares and cross-products of the original explanatory variables
- This model is then tested for validity using an F-test, or as a shortcut method by using the fact that if the null is true that all the parameters in the regression are equal to zero, then the F-stat should be distributed according to the rule: $\chi^2 = N * R^2 \sim \chi^2_{S-1}$, 214

Conducting the White Test

The Models:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + e_i$$
$$z_2 = x_{2i},\ \ z_3 = x_{3i},\ \ z_4 = x_{2i}^2,\ \ z_5 = x_{3i}^2,\ \ z_6 = x_{2i}x_{3i}$$
$$e_i^2 = \delta_1 + \delta_2 z_{2i} + \delta_3 z_{3i} + \delta_4 z_{4i} + \delta_5 z_{5i} + \delta_6 z_{6i} + \text{error}$$

Hypotheses:

$$H_0: \delta_2 = \delta_3 = \delta_4 = \delta_5 = \delta_6 = 0$$
$$H_1: \text{at least one of } \delta_i \neq 0$$

The test-statistic is given by:

$$\chi^2 = NR^2 \sim \chi^2_{s-1}$$

Where *N*=sample size, R^2=coefficient of determination, s=number of variables in e_i^2 model

The larger the result of the χ^2 test statistic, the less likely it is that the null is true, and hence the more likely it is that the errors are not normally distributed. Note: do not include cross products for triple variable combinations or squared terms and non-squared terms.

Assumption Four: Nonstationarity I

Stationary and Nonstationary Variables

- A time series is stationary if its mean and variance are constant over time, and if the covariance between any two particular values does not vary systematically with time, 326
- Any time series that either has a clear trend, or has a wandering 'random walk' pattern over time, is said to be nonstationary, 327-328

Spurious Regressions

- It is important to know about nonstationarity because if we regress two or more nonstationary variables against each other, we may get a relationship that is falsely highly significant, with an inflated R^2, 333
- This occurs essentially because both series are changing in a similar way over time, even though this behaviour has nothing to do with the behaviour of the other series, 333-334
- Nonstationarity thus produces incorrect results for t-statistics, 334

Lagged Independent Variables

- In many cases we must allow for the possibility that time might elapse between a change in the independent variable and the resulting change in the dependent variable
- Many econometric equations include one or more lagged independent variables like x_{t-1} where the subscript $t-1$ indicates that the observation of *x* is from the time period previous to time period *t*
- The simplest dynamic model is an equation in which the current value of the dependent variable *y* is a function of the current value of *x* and a lagged value of *y* itself
- For example: $y_t = \alpha_0 + \beta_0 x_t + \lambda y_{t-1} + e_t t, \;\; 0 < \lambda < 1$

The First Difference

- In dealing with nonstationarity, it is useful to consider the concept of the first difference
- The first difference is simply the difference between any given y value and its preceding y value, 326
- The first-order autoregressive model refers to a simple regression model where each y value is regressed against the previous value of itself (y-1), along with a random error term, 329
- For example: $y_t = \rho y_{t-1} + v_t$

Unit Roots

- Note that if the coefficient in front of the y_{t-1} term (ρ) is less than one, it can be shown that the expected value of y_t for sufficiently large samples is zero, 329
- This means that such a model fluctuates around a mean of zero
- We can have the model instead fluctuate around some number aside from zero by including an intercept term α into the autoregression model, 329-330
- If we want a model that fluctuates about some progressive trend, we can include an additional trend term λt, 331

Random Walk Models

- These occur in the special case where $\rho = 1$, 331
- They exhibit unpredictable wandering behaviour, which is clearly not a trend but not mere random fluctuations either, 331
- The reason for this behaviour is that the unit coefficient permits past errors to built upon themselves and add together into temporary 'trends', 331
- This does not occur if $\rho < 1$ because in this case the influence of past errors will be multiplied by ρ^t, which being a decimal will quickly reduce to zero and hence past errors will have little effect on present values of y, 331
- If we include a constant term into a random walk model, we get what is called a random walk with drift, as the random walk now tends to trend up or down, with wandering behaviour around this trend, 332
- This occurs because successive intercept values are added together to reinforce each other, just as occurs for the error terms, 332
- Inclusion of a specific trend term into a random walk model will emphasise the trending behaviour even further, 333
- It can be shown that all random-walk models of the types discussed here have nonstationary variance, and often nonstationary mean, 333

The Dickey-Fuller Test

- The Dickey-Fuller test is the most popular method for testing to see if a series is stationary or not, 335
- To do this, we simply construct a hypothesis test with a null that the ρ term in $y_t = \rho y_{t-1} + v_t$ equals one; if so, it means the series is nonstationary, 335
- Note that we can also conduct dickey-fuller tests on series with constants and trends, 335
- If ρ=1 and hence the series is nonstationary, we say that the series has a unit root, 337
- Note, however, that because of the varying standard deviation in the case of nonstationarity under the null, the variable will not be t-distributed, but rather will follow a special distribution called the τ-distribution, which has its own special critical values, 336

Conducting the Dickey-Fuller Test

The Models:

$$y = \rho y_{t-1} + v_t$$
$$y - y_{t-1} = \rho y_{t-1} - y_{t-1} + v_t$$
$$\Delta y_t = y_{t-1}(\rho - 1) + v_t$$
$$\Delta y_t = \gamma y_{t-1} + v_t$$

$$\text{where } \gamma = (\rho - 1)$$

$$\Delta y_t = \gamma y_{t-1} + v_t$$
$$\Delta y_t = \alpha + \gamma y_{t-1} + v_t$$
$$\Delta y_t = \alpha + \gamma y_{t-1} + \lambda t + v_t$$

Hypotheses:

$$H_0: \gamma = 0$$
$$H_1: \gamma < 0$$

The test-statistic is not given: will need to be provided with critical values

The null states that it is a nonstationary process (ρ=1), so if we reject the null we conclude that it is a stationary model.

Assumption Four: Nonstationarity II

Introduction to Cointegration

- As a general rule nonstationary time-series variables should not be used in regression models, to avoid the problem of spurious regressions
- However, there is an exception to this rule – the case of cointegration
- Cointegration It is a way of matching the degree of nonstationarity of the variables in an equation in a way that makes the *error term* of the equation stationary and rids the equation of any spurious regression results
- Even though individual variables might be nonstationary it's possible for linear combinations of nonstationary variables to be stationary or *cointegrated*
- If a long-run equilibrium relationship exists between a set of variables those variables are said to be *cointegrated*

Benefits of Cointegration

- Although two series might display random walk behaviour, it is possible that the two trends may 'wander together', 339
- If this is the case, it means that the two series are never very far apart for very long, and hence a new equation that is a linear combination of the two series may itself by stationary, 339
- We can test for this by running a τ-test on this new equation, though in this case the critical values will once again be different as this time we are using estimated errors (residuals) to test rather than actual observed y values, 339
- In this case, the null hypothesis will be that the series are not cointegrated, and hence the residuals are nonstationary, 340

Order of Integration

- The order of integration of a series is the minimum number of times it must be differenced to make it stationary
- Stationary series are said to be integrated of order 0: I(0)
- A series that is nonstationary but its first difference $\Delta y_t = y_t - y_{t-1}$ is stationary is said to be integrated of order I(1)

- We can use the Dickey-Fuller test to determine if the first or second integrated order of a regression is stationary or not

Testing for Cointegration

- The test for cointegration is effectively a test of the stationarity of the least squares residuals
- The standard *t*-tests do not apply to this application so adjusted critical *t*-values should be used; this means that we cannot use the standard p-values put out be Eviews
- Also note that the regression has no constant term because the mean of the regression residuals is zero

Conducting the Cointegration Test

The Models:

$$\hat{y}_i = b_1 + b_2 x_{2i} + \hat{e}_i$$
$$\Delta\hat{e}_t = \gamma\hat{e}_{t-1} + v_t$$

Hypotheses:

H_0: the series are not cointegrated ⇔ the residuals are nonstationary
H_1: the series are cointegrated ⇔ the residuals are stationary

The test-statistic is not given: will need to be provided with critical values

The null states that the residual is nonstationary, so if we reject the null we conclude that residuals are stationary (constant), and hence the two variables in the original equation are cointegrated.

Procedure for Dealing with Nonstationary Series

- 1. Specify the Model
- 2. Test all variables for nonstationarity using the appropriate version of the Dickey-Fuller test
- 3. If the variables don't have unit roots, estimate the equation in its original units (*y* and *x*)
- 4. If the variables have unit roots and if the dependent and explanatory variables have unit roots (i.e. I(1)), test the residuals of the equation for cointegration
- 5. If the variables have unit roots but are not cointegrated then change the functional form of the model to first differences (Δy and Δx) and estimate the equation
- 6. If the variables have unit roots and also are cointegrated then estimate the equation in its original form

Assumption Five: Multicollinearity

What is Collinearity?

- When data are the result of an uncontrolled experiment (as is often the case in econometric work), it is possible that many of the variables will move together in systematic ways, 153
- This problem is referred to as collinearity, 153
- It will be difficult to distinguish the separate effects from the two variables if they are so closely correlated in this way, 153

Exact Multicollinearity

- If one variable is an exact linear function of another, it will be impossible to determine the values of the coefficients themselves
- All we will be able to get are the relationships between the two variables (e.g. sum or difference between them)
- The dummy variable trap is an example of collinearity, as if $x_2 + x_3 = 1$, then we can substitute x_2 for $x_2 = 1 - x_3$
- The new regression equation could be written as:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3$$
$$y = \beta_1 + \beta_2 (1 - x_3) + \beta_3 x_3$$
$$y = \beta_1 + \beta_2 - \beta_2 x_3 + \beta_3 x_3$$
$$y = (\beta_1 + \beta_2) + (-\beta_2 + \beta_3) x_3$$

- Hence we can see there is only enough information to estimate two coefficients and not three coefficients
- In Eviews if you try to estimate a model in which exact multicollinearity is present an error message of the form "*near singular matrix*" will be given

Imperfect Multicollinearity

- Imperfect multicollinearity occurs when two (or more) explanatory variables are imperfectly linearly related
- For example, $X_{1i} = \alpha_1 + \alpha_2 X_{2i} + v_i$
- Notice that this equation includes vi a stochastic error term
- This implies that while the relationship between X1 and X2 might be fairly strong, it is not strong enough to allow X1 to be completely explained by X2; some unexplained variation remains
- Imperfect multicollinearity could be a fluke of the sample, or it could be an underlying theoretical property of our data

Consequences of Collinearity

- 1. If collinearity is high but not perfect, estimation of all regression coefficients is still possible, and the estimates remain BLUE – that is they are unbiased and efficient
- 2. The major consequence of multicollinearity is that the variances of the estimates will increase; since two or more of the explanatory variables are significantly related it becomes difficult to precisely identify their separate effects
- 3. Multicollinearity increases the variance, estimated variance and therefore the standard error of the estimated coefficient. Hence, t-statistics will be reduced, making it more likely that we will be unable to reject the null hypothesis that the coefficient has a zero value
- 4. Estimates will become very sensitive to changes in model specification, meaning that the addition or deletion of an explanatory variable or a few observations will often cause major changes in the values of the estimated coefficients. These large changes occur because OLS estimation is sometimes forced to emphasize very small differences between variables in order to distinguish the effect of one multicollinear variable from another.
- 5. The overall fit of the equation will be largely unaffected. Even though individual t-statistics are often quite low, the overall fit of the equation as measured by R2 or the F-test will not fall.

Detection of Multicollinearity

- Since multicollinearity is essentially a sample phenomenon there is not one unique method of detecting it or measuring its strength. Nonetheless, common signs include:
- 1. High R^2 and F-tests with low t-statistics
- 2. High Values for correlation coefficients (though this only works for relationships between two explanatory variables)
- 3. Regression coefficients are highly sensitive to model specification
- 4. Another method is to use an auxiliary regression, where one of the explanatory variables is treated as the dependent variable, and the resultant R^2 is examined; if it is high, then most of the variation in that variable is explained by the variation of other variables, an indication of collinearity, 156

Solutions to Multicollinearity

- When we have a well-specified model with highly significant parameters of the expected sign and magnitude, there is no need to worry about collinearity, as by itself it does not violate any OLS assumptions, 155
- 1. Eliminating Variables: We have to be careful though that we don't drop an important explanatory variable from the equation and introduce specification bias (hence this is usually a last resort)
- 2. Reformulating the Model: For example by forming a linear combination of the multicollinear variables or by transforming the equation using first differences or logs
- 3. Additional or new data: For example, increasing the sample size may alleviate the problem since a larger data set will allow more accurate estimates than a smaller one (though it is often hard to get additional data)
- 4. Re-examine Purpose: We try to avoid multicollinearity by choosing variables that are theoretically relevant (for meaningful interpretation) and that are also statistically non-multicollinear (for meaningful inference)
- 5. Inclusion of non-sample data to alter the model

Nonsample Information

- Sometimes we can use additional information not included in our sample in order to construct more accurate or refined estimates, 146
- For example, we might expect from economic theory that the demand for beer would depend upon income, the price of beer and the price of other alcoholic products, but if the price of these all went up by the same proportion, we would expect demand not to change
- By substituting a constant in front of each regression parameter and simplifying to bring all these terms together (see algebra on page), we find that we can place a restriction on the allowable parameters, such that $\beta_1 + \beta_2 + \beta_3 = 0$
- We can then use this information to place an additional restriction on one of the parameters: e.g. $\beta_4 = -\beta_2 - \beta_1$
- Substituting this into the original regression, we find that we have eliminated a parameter, thereby reducing the variance (spread) of our resulting estimates, 147-148
- Note, however, that these estimates will only remain unbiased if the restriction we introduced is exactly true, 148

Assumption Six: Non-Normal Errors

The Benefit of Normally Distributed Errors

- Construction of tests and confidence intervals relies on the assumption that errors are normally distributed, 89
- Although for large sample sizes this assumption is not crucial, it is still useful to know whether or not it holds, 89
- For example, when choosing a functional form one of the criteria that could be used is whether it leads to errors that are normally distributed

The Jarque-Bera Test

- The Jarque-Bera test is used to test the assumption of normally distributed errors
- The Jaque-Bera test relies on use of the skewness and kurtosis values of the residual distribution to test whether the residuals are likely to be normally distributed, 89
- For a normal distribution the kurtosis value is 3 and skewness is 0
- When the residuals are normally distributed, the Jaque-Bera test has a chi-square distribution with two degrees of freedom, 90
- Thus, we can test for whether or not the JB statistic is likely to be chi-square distributed, and hence infer whether our residuals are likely to be normally distributed, 90
- We reject the hypothesis that the JB statistic has a chi-square distribution if the JB stat takes on values larger than certain defined critical values, which will depend upon the chosen level of significance, 90

Conducting the Test

The Model:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + e_i$$

Hypotheses:

H_0: normally distributed disturbances
H_1: non-normally distributed disturbances

The test-statistic is given by:

$$JB = \frac{N}{6}\left(S^2 + \frac{(k-3)^2}{4}\right) \sim \chi^2$$

Where *N*=sample size, *S*=skewness and *K*=kurtosis

The larger the result of the JB test statistic, the less likely it is that our errors are normally distributed.

Part F: Exam Revision Stuff

Tutorial Questions to Check

- The estimates and t-statistics associated with the other explanatory variables should not change much if we remove an irrelevant variable from the model
- Summation operators are not affected by differentiation
- Absolute value comparisons of t-statistics only work for two-tailed tests
- To find *SST* we use $\hat{\sigma}_y = \sqrt{\frac{SST}{N-1}}$
- If a generalised/weighted least squares estimator has a lower variance than the regular or white version, this indicates that the weighting assumption was valid
- Know what a chow test is – structural change dummy F-test
- Assumptions are always stated in terms of the error terms, never in terms of the residuals
- The dickey-fuller test is a modified t-test, with $H_A: \gamma < 0$, null=non-stationary
- If there is a genuine relationship between two variables, a strong regression relationship should hold even after taking the first difference
- When asked to comment about regression results, make sure you mention the signs of the explanatory variables and what these indicate
- Use correct formula for interpretation of log dummy variables!
- Standard Error of Regression $= \sqrt{\hat{\sigma}^2}$
- When interpreting a quadratic model, it is a good idea to work out the turning point
- $\Delta \log(y) = \beta_2 \Delta x \text{ then } \%\Delta y \cong 100\beta_2 \Delta x$
- H_0: the series are not cointegrated ⇔ the residuals are nonstationary
- H_1: the series are cointegrated ⇔ the residuals are stationary