ECONOMETRIC MODELLING AND APPLICATIONS

Prof. Dr. Aysu İnsel

Marmara University

Department of Economics

2009-2010 ***Econometrics Lecture Notes***

http://mimoza.marmara.edu.tr/~ainsel/
ainsel@marmara.edu.tr

CONTENT

1-Statistical Review:

1.1 Random Variable:

A random variable (r.v.) is a variable whose value is unknown until it is observed. The value of a random variable is obtained from an experiment and it is not perfectly predictable. A random variable is a function of the sample space.

- A random variable is called discrete if it can take only a finite number of values and they can be counted by using the positive integers.
- A random variable is called continuous if it can take any real value in an interval on the real number line.

Each random variable can be described using the probability density function (pdf) or the cumulative density function (cdf):

- For ***a discrete random variable X***, the value of the probability density function f(x) is the probability that the random variable X takes the value of x, f(x)=P(X=x).

- For ***a continuous random variable*** Y, the probability density function f(y) can be calculated by an equation.

1.2 Random Process:

The theory of random process was developed in order to explain the fluctuations and noise in engineering studies. However, a sample data on different economic variables is also a result of a random process.

A random process is a collection of random variables defined on a given probability space.

Random process is described by using the statistical expectations, covariance, variance, and correlation functions.

1.3 Expectation, Covariance, Variance, and Correlation:

The ***expected value*** of a r.v. X is the average of the random variable in repeated samples, and it is defined by

$$E[X] = \sum_{i=1}^{n} x_i \, f(x_i) = \mu \qquad \text{if it is discrete r.v.}$$

$$E[X] = \int_{-\infty}^{+\infty} x \, f(x) \, dx = \mu \qquad \text{if it is continuous r.v.}$$

The ***variance*** of a r.v. X shows the variability of the random variable in repeated samples, and it is defined by

$$V[X] = E[X - E(X)]^2 = E\left[X^2\right] - [E[X]]^2 = \sigma^2$$

$$= \sum_{i=1}^{n} (x_i - \mu)^2 \; f(x_i) \qquad \text{if X is discrete r.v.}$$

$$= \int_{-\infty}^{+\infty} (x - \mu)^2 \; f(x)\, dx \qquad \text{if X is continuous r.v.}$$

The ***covariance*** of two r.v.s X and Y represents the direction of co-movements. In other words, it measures the tendency of X and Y moving in the same direction or in the opposite direction.

Let f(x,y) is the joint pdf of random variables X and Y. Then, covariance is defined by

$$Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)] = \sigma_{XY}$$

$$= \sum_x \sum_y [x - E(X)][y - E(Y)]\, f(x,y) \quad \text{if X and Y are discrete r.v.}$$

$$= \int_x \int_y (x - E(X))(y - E(y))\, f(x,y) \quad \text{if X and Y are continuous r.v.}$$

If $\sigma_{XY} > 0$, X and Y move in the same direction.
If $\sigma_{XY} < 0$, X and Y move in the opposite direction.
If $\sigma_{XY} = 0$, X and Y are independent.

If X and Y are ***independent***, then covariance will be zero. However, zero covariance does not imply an independence of variables X and Y.

The ***correlation*** of two random variables X and Y measures the strength of linear relationship between them. It is defined by

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)\, V(Y)}} = \rho$$

If ρ =1, X and Y have a strong positive correlation.

If ρ=-1, X and Y have a strong negative correlation.

1.4 Useful Distributions:

Statistical inference process, such as estimation and testing, heavily depends on distributions. Some important distributions are as follows:

- ***Normal Distribution***

If random variable X is a normal variable, X~N(μ,σ^2), then its distribution function is

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{1}{2\sigma^2}(x-\mu)^2).$$

- ***Standard Normal Distribution***

If the random variable Z is a standard normal variable, Z~N$(0,1)$, with the zero mean and unit variance, then its distribution function is

$$f(z)=\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}(z)^2)\ .$$

- ***Chi-Square Distribution***

Chi-Sq distribution with degrees of freedom (d.o.f.) m is defined as the sum of n independent, squared standard normal variables. If an independent standard normal variable Z has a zero mean and unit variance, i.e. Z~(0,1), then

$$Y=\sum_{i=1}^{n}z_i^2=\chi^2(m)\ .$$

The degrees of freedom can be considered as the number of independent information.

- ***Student t Distribution***

Student t distribution is defined as the ratio of a standard normal variable to the square root of an independent chi-squared variable divided by its degrees of freedom m provided Z and Y are independent. That is $\frac{N(0,1)}{\sqrt{\frac{\chi^2}{m}}}=\frac{Z}{\sqrt{\frac{Y}{m}}}$ ~ t(m).

- ***F Distribution***

If Y and W are two independent chi-squared random variables with degrees of freedom m and k respectively, then the ratio of two independent chi-squared random variables each divided by their degrees of freedom is distributed as the F distribution with d.o.f. m and k. That is $\frac{Y/m}{W/k}$ ~ F(m,k).

1.5 Class of Random Process:

A collection of random variables ordered in time instants is called a ***stochastic*** or ***random process***.

The probability structure of the sequence of random variable X(t) is determined by the joint distribution of a stochastic process.

Thus random (stochastic) process can be characterized entirely by the mean and variance. A random (stochastic) process can be classified as follows:

- ***Stationary Stochastic Process***:

A random process {X(t), t∈T} with mean μ and variance σ^2 is said to be ***stationary***, *if its mean and variance are independent of time instants* $\{t_i \in T, i=1,2,..,n\}$ *for all n.*

In other words, the distribution of a stationary variable X will be unchanged when changing the time period or rearranging the ordering of the time index. That is, it will have the same mean and variance for every time period and they do not depend on the time but on the time lag.

So, **stationarity implies a restriction on the time heterogeneity of the process.**

- ***Independent Stochastic Process:***

In a random process X(t), if $X(t_i)$ for all i=1,2,..,n are independent random variables, then X(t) is called an ***independent process***.

So, **independence implies a restriction on the memory of process.**

- ***Independent Stationary Stochastic Process:***

In a random process X(t), if $X(t_i)$ are independent for all i=1,2,..,n and *has the same joint distribution*, then it is called an ***independent stationary process***.

- ***Markov Process:***

Markov process states that the future state of the process is determined only by the present state, not by the past history.

- ***Normal (Gaussian) Process:***

A random process is said to be a ***normal process*** if X(t) follows a normal distribution and also its distribution is completely characterized by the mean and variance.

If the joint distribution of $X(t_i)$ is the same as the joint distribution of $X(t_{i+\Gamma})$ for all i=1,2,..,n and Γ, then the process is ***purely random process***.

In other words, the distribution of the process remains unchanged when the time is shifted by an arbitrary value of Γ. This process is also known as the ***strictly stationary process***.

- ***Ergodic Process:***

A random process is called ***ergodic process***, when not only the mean and variance but also the **whole distribution unaffected** by a change in time horizon.

An independent stationary stochastic process is ergodic and has a limited memory.

1.5 Data Analysis:

1. Data Transformation, and Decomposition:

In the most empirical studies, the raw data is re-defined and/or adjusted using different transformation techniques.

For example:

- ☞ the level series on the nominal foreign exchange rate (TL/$) and on the consumer price index are converted into annual (year to year) or monthly (month to month) percentage changes, and
- ☞ each series may be decomposed into seasonal, cyclical, trend, and irregular components by the X-12-ARIMA or Tramo/Seats methods and then, the HP filter is used to obtain the Trend and Cycle components separately.

Economic time series generally exhibits

- the regular seasonal peaks and troughs related to the calendar effects (seasonal component),
- the long term movements or the direction of the data over time (trend component),
- the short term oscillations (cyclical component) and
- the short term neither systematic nor predictable random fluctuations (irregular component).

Since seasonal adjustment procedure removes seasonal fluctuations, the resulting series is much smoother than the original series.

The seasonally adjusted series **consists of** the **trend-cycle and the irregular components.**

If the amplitudes of seasonal and irregular components **do not change** as the trend components changes, an additive decomposition model is the appropriate one.

However, if their amplitudes **change**, then multiplicative decomposition model will be the appropriate one.

The trend-cycle and seasonally adjusted components **should be consistent**.

The original series is equal to the sum of the seasonally adjusted series and the seasonal factors. Since the seasonal components consists of reasonably systematic and stable effects with respect to timing, direction and magnitude, these effects over time are captured by the seasonal factor.

Seasonal factors are generally calculated depending upon the patterns of the seasonal fluctuations in past years and upon the unknown pattern of seasonal fluctuations in future years.

In general, if the time series defined in terms of annual changes, the **trend component** exerts a strong influence on the series, but the magnitude of the seasonal component becomes much smaller relative to monthly or quarterly changes.

Note first that, from the econometric perspective, the trend-cycle and seasonally adjusted series could appear as complements since they both are independent of regular variations and they can reflect real economic movements better than original series.

Second, from the economic perspective, information on seasonal factors can enable policy makers to distinguish the seasonal changes and the long run changes in economic variables, and hence they can design appropriate policy responses.

2. Graphical Analysis:

While transforming and decomposing time series can facilitate a more efficient statistical analysis, construction and description of figures and/or scatter diagrams can provide a visual inspection of the time series.

If possible, different lengths of the whole sample period can be selected for variables of interest in order to cover the economic crises and different policy applications. In other terms, a sub-period analysis provides more information about the reasons and consequences of different economic policy applications.

3. Descriptive Analysis:

Visual inspection can be extended to the descriptive statistical analysis using the mean, the standard deviation, the skewness, the kurtosis, and the maximum/minimum values in order to provide a powerful statistical data analysis.

It is believed that most of the economic time series do not fulfil the independent Gaussian distribution. A common problem in the economic time series is the presence of persistence.

In general, a tendency for large (small) values is to be followed by large (small) values. For example, interest rate or exchange rate series exhibit persistence because these variables are usually used as the policy instruments and thus their behaviour cannot be arbitrary. So, it is apparent that data performances of these variables are crucial in economic modelling and forecasting.

4. Correlation Analysis:

Simple correlation and cross correlation coefficients are more accurate and comprehensive quantitative measures in assessing the strength, the timing, and the direction of the relationship between two variables.

Linear nature of the contemporaneous relationship between two random variables in (different) sample periods can be described by the simple **correlation coefficient**, whereas a phase relationship between two variables can be classified by the **cross correlation coefficient**. Thus the direction and the time patterns of variables can be determined in order to provide some clues/signals about the policy impacts.

5. Causality Analysis:

Existence of some causal relationship between two economic variables indicates a non-contemporaneous relationship between them. The **Granger causality test** provides a basis for addressing a feedback effect of one variable on another.

6. Range from data description to econometric methods:

Time series econometric methods play important role in empirical studies, (Diebold-Introduction to time series).

Nonstructural econometrics influences theory by

- ☞ helping to identify the facts that need explanation

Theory influences non-structural econometrics by

- ☞ guiding the choice of series examined.

Structural econometrics

- ☞ influenced by theory through testable restrictions, BUT
- ☞ affect theory by helping separate good theory from bad theory.

Non-structural (Modelling) Econometrics:

Data description: To get idea in a systematic way in order to provide information for

- ☞ the theory
- ☞ (serial) correlation pattern
- ☞ Persistence
- ☞ Volatility

Since economic variables are observed over time, and interested in

- ☞ relationship between the current & past values of a variable (univariate case);
- ☞ interrelationship between/among current & past values of variables (bivariate or multivariate case).

Difference between the **time series analysis and the multivariate analysis** is the temporal pattern of the observed data.

Questions about temporal pattern of data:

1) Is the series varying about a fixed level in a **stable way**?
2) Is the **level changing**?
3) Is there a **trend**?
4) Is the **trend changing**?
5) Is there a **seasonal pattern**?
6) Is the seasonal pattern **changing** over time?
7) Is there any **secular variation**?
8) Is there any **cyclical variation**?
9) Is there any **exogenous shocks/intervention**?
10) Is there any **past history**?

Short Answers:

1) Description of Trend and Detrending:
   - ☞ Stochastic trend
   - ☞ Deterministic trend
   - ☞ Hodrick-Prescott (HP) filter
   - ☞ Perron detrending
2) Description of Seasonality and Seasonal Adjustment:
   - ☞ Deterministic seasonality
   - ☞ Stochastic seasonality
3) Description of Cycles:
   - ☞ Business cycle
4) Causality, Exogeneity:
   - ☞ Direction of causality-which variable causes the other one to move. Granger causality
   - ☞ Feedback effects
   - ☞ Exogeneity: weak ex. (need to estimate and testing); strong ex. (need to forecast); super ex. (need policy analysis).
5) Unit roots, Persistence, Cointegration.

Structural (Modelling) Econometrics:

- ☞ Structural microeconometrics
- ☞ Structural macroeconometrics

Applications of

- ☞ Time series models (ARMA), and
- ☞ Distributed Lag or Autoregressive Distributed Lag Models (ARDL).

2- Model Specification, Estimation and Testing:

Regression analysis enables us to characterize the stochastic relationship between a variable of interest called ***dependent, endogenous***, or ***response variable*** and one or more related ***independent, explanatory, exogenous***, or ***predictor variables***.

Since econometrics is a methodology for estimating different economic relationships, **economic theory** provides the starting point by identifying the related variables, determining the behavior of variables, and stating the direction of causality.

A relation between two random variables X and Y can be specified as a ***functional relation*** and ***statistical relation***.

Functional relation between X and Y is *exact* and the value of Y is *uniquely* determined by the value of a specified X. It is a **mathematical model** for the response of Y.

Hence all the observations fall on the line of relationship.

On the other hand, **statistical relation** between X and Y is *not exact* and the value of Y is *not uniquely* determined by the value of a specified X. It is a **deterministic mathematical model** and it enables us to estimate Y for various values of X but it does not provide a way in order to evaluate the estimation error. However, most of the observations deviate from the line of relationship and display a random pattern.

A method of fitting a model to data must provide a bound on the errors of estimating Y for given values of X.

The ***probabilistic model*** is a simple modification of the deterministic model by considering the conditional mean of Y for given value of X and the random error.

For example, consider a linear demand equation,

$$Q = \alpha + \beta P \qquad (1)$$

Economic theory postulates an inverse relationship between the quantity (Q) and price (P).

This is a ***bivariate case***. A ***bivariate probability density function*** of quantity values exits for each price.

It must be noted that there are some difficulties in modeling the relationship between Q and P such that when the actual data on Q and P are used for estimating equation (1), a range of quantity values will be observed for each price. The values will not be equally likely, and thus a probability density function of Q will exist for each price.

If the demand function is **stochastic**, higher prices are not necessarily related to the lower quantities demanded.

The relationship between Q and P can be explained in terms of the **conditional mean** of quantity demanded as a function of the price of a good.

The conditional mean of Q for a given value of P is denoted $\mathbf{E(Q \mid P_i)}$, AND

The **empirical model** for a particular observed value of Q is equal to **the conditional mean of Q for a given value of P plus a random error, i.e.** $Q_i=\alpha + \beta P_i + \varepsilon_i$ **(2)**

So, Q varies in a random manner and has a conditional distribution with a mean,
$E(Q \mid P_i) = \alpha + \beta P_i$,
implying that **variations in Q** can be explained in terms of
(a) The variations in the mean, and
(b) The variations around the mean.

The conditional mean function is the *regression function*. **Regression function** relates the actual quantity to price and incorporates the statistical relation by assuming that
(a) For each level of P, there is a probability distribution of Q, and
(b) The mean of the probability distribution varies in a systematic way with P.

If the statistical relation in equation (2) is assumed to be linear, then the regression function,
$(\alpha + \beta P_i)$,
is the **deterministic** or **systematic part** of the ***regression component*** and it is the ***line of statistical relationship component.***

The **indeterministic** or **unsystematic part** is the **stochastic error** or **random disturbance component**, (ε_i).

2.1 Classical Linear Regression Models:

Consider a **simple linear regression model**

$$Y_i=\alpha + \beta X_i + \varepsilon_i \qquad (3)$$

and the assumptions about the random error term ε_i and Y_i and X_i are :

1) The error term is normally distributed,
2) The expected value of the error term is zero, i.e. $E(\varepsilon_i)=0$,
3) The variance of the error term is constant at all levels of X, $V(\varepsilon_i)= \sigma^2$ and error terms are independent of each other, $E(\varepsilon_i, \varepsilon_j)=0$,
4) The error term is statistically independent, i.e. $\varepsilon_i \sim$**i.i.d.**$(0, \sigma^2)$.

   Assumption 2 shows zero mean and ensures the ***unbiasedness*** of the estimators a, b, and S^2. Assumption 3 implies ***homoscedasticity*** and ***nonautocorrelation***.
   Assumption 4 states that error terms are ***identically and independently distributed,*** and hence they are generated from a *stationary random process*.
5) The model is ***linear in parameters***, and both α, β are fixed parameters. Parameter α is the intercept term and β is the slope term.
6) Explanatory variable X is ***non-stochastic***, in other terms x_i is fixed.

Simple linear regression model has three parameters to be estimated, namely α, β, and σ^2.

Assumptions are the basic elements of a simple regression model in order to develop a structure for the model prediction, and to provide insights about how the economy operates.

The regression model can be considered as an ***empirical model***, and the model must be consistent with the manner in which the economic variables actually behave.

It is worth noting that **the mean of Y** represents the true regression line, and **the variance of Y** for fixed x shows the ***variability*** of Y at a particular value of x.

The variability in Y is determined by the error variance σ^2. That is

The mean of Y: $E[Y]= E[\alpha + \beta X + \varepsilon]$

$$= \alpha + \beta X + E[\varepsilon] = \alpha + \beta X \qquad (4)$$

The variance of Y: $V[Y]= V[\alpha + \beta X + \varepsilon]$

$$= \alpha + \beta X + V[\varepsilon] = \sigma^2 \qquad (5)$$

Since $E(\varepsilon) = 0$ and $V(\alpha) = V(\beta) = 0$, as α and β are fixed parameters.

A. Properties of a Good Estimator:

A good estimator must be: ***(a) unbiased, (b) efficient,*** and ***(c) consistent***.

(a) Unbiasedness: If the expected value of an estimator equals the true parameter, **E[** $\hat{\theta}$ **]=** θ, then the estimator is **unbiased,** otherwise the estimator is called to be **biased**.

That is, an estimator, at least on average, should be equal to the parameter that is supposed to be estimated.

(b) Efficiency: If an unbiased estimator has the ***smallest possible variance***, then it is **efficient**.

$$\hat{\theta}_1 \text{ and } \hat{\theta}_2 \text{ are two unbiased estimators,}$$

That is, if $\text{and } V(\hat{\theta}_1) \rangle V(\hat{\theta}_2)$

then $\hat{\theta}_2$ is ***more efficient*** estimator.

(c) Consistency: Probability limit of an estimator approaches to the true parameter as the number of observations increases. **Consistency** is an asymptotic property.

In other words, if $\lim_{n\to\infty} P(\left|\hat{\theta} - \theta\right| \langle c) = 1$, for c>0, then the estimator is said to be **consistent**.

B. Least Squares Principle:

As discussed above, an empirical model must estimate or predict the economic behaviors successfully, and then it must provide correct information about how the economy operates in the real state.

There are many methods for estimation of an empirical model. However, the ***ordinary least squares (OLS)*** approach is the most commonly used one as long as the above six assumptions are satisfied by a linear regression model.

Consider any linear combination of Y and X, given by $y_i=\alpha + \beta x_i$, the *least squares principle* is to ***minimize*** the squared deviation of y_i from $a+bx_i$, where **a** and **b** are estimators of **α** and **β**.

In other terms, ***the least squares principle*** is ***to choose the estimators a and b so that the sum of squared differences must be the smallest one***.

That is

$$L = \min \sum_{i=1}^{n}(y_i - a - \beta x_i) = \sum_{i=1}^{n}\varepsilon_i^2 \Rightarrow \text{Using the first order condition}$$

$$\frac{\partial L}{\partial a} = -2\sum_{i=1}^{n}(y_i - a - bx_i) = 0 \Rightarrow \bar{y} - a - b\bar{x} = 0$$

$$\frac{\partial L}{\partial \beta} = -2\sum_{i=1}^{n}(y_i - a - bx_i).x_i = 0 \Rightarrow \sum_{i=1}^{n} y_i x_i - a\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^2 = 0 \qquad (7)$$

Equations in (6) and (7) are called ***normal equations***.

They can also be written in terms of **OLS residuals** $\sum_{i=1}^{n}\hat{\varepsilon}_i = 0$ and $\sum_{i=1}^{n} x_i\hat{\varepsilon}_i = 0.$

These equations simplify that the residuals and explanatory variables are orthogonal to each other, and *the least squares residual sum to zero* under the presence of the intercept term.

The intercept term (a) and the slope term (b) are estimated by the following equations

$$a = \bar{y} - b\,\bar{x}$$

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\text{sample Cov}(x, y)}{\text{sample Var}(x)}$$

C. *Distribution of Estimators:*

Consider a linear regression function of Y given X such as $E[Y_i] = \alpha + \beta X_i$ for unknown constant parameters α and β.

The error between Y and its average is an identically, independently and normally distributed random variable with distribution N(0, σ^2).

Thus the regression function can be written as $E[Y_i] + \varepsilon_i = \alpha + \beta X_i + \varepsilon_i$.

The estimators for the true intercept term is denoted by a, and for the true slope term is denoted by b.

The least squares estimators of the regression coefficients are random variables since their values depend upon the observed values of Y. The distribution of least squares estimators can be viewed as follows:

$$a \sim N\left(\alpha,\ \sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)\right)$$

$$b \sim N\left(\beta,\ \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)=N\left(\frac{\sigma^2}{(n-1)\,\text{sample Var}(X)}\right)$$

Thus the estimators a and b are both **unbiased** and also they are **efficient.**

The ***Gauss-MarkovTheorem*** states that the OLS estimators are the *best linear unbiased estimators* ***(BLUE)***.

Finally, the third estimator in the simple regression model is the estimator of σ^2, denoted by S^2. The estimator S^2 is obtained from

$$s^2=\frac{1}{n-2}\sum_{i=1}^{n}(y_i-a-bx_i)^2=\frac{1}{n-2}(y_i-\hat{y}_i)^2=\frac{1}{n-2}\sum_{i=1}^{n}\hat{\varepsilon}_i$$

$\hat{y}_i = a + bx_i$ gives the fitted values , and

where $\hat{\varepsilon}_i = y_i - a - bx_i$ gives the OLS residuals

The estimator S^2 has a Chi-squared distribution with d.o.f (n-2), i.e.

$$\frac{n-2}{\sigma^2}s^2 \sim \chi^2_{n-2}.$$

Multiple Linear Regression Models:

Consider the multi-variables classical regression model

$y = X\beta + \varepsilon,$

where **y** denotes an **(nx1)** vector of dependent variable and **X** denotes an **(nxk)** matrix of observations on explanatory variables, **β** is a (kx1) vector of unknown (constant) parameters and **ε** denotes an **(nx1)** vector of observations on an unobservable disturbance term.

THE LINEAR REGRESSION MODEL IN MATRIX NOTATION

In ordinary algebra, the k-variable linear regression model is

$$Y_t = \beta_1 + \beta_2 X_{t2} + \ldots + \beta_k X_{tk} + u_t \quad , t = 1, \ldots, T \qquad (1)$$

Define a k×1 vector of parameters

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_k \end{bmatrix}$$

and then, a 1×k row vector of the t [th] observation on each of the k variables:

$$x_t = \begin{pmatrix} 1 & X_{t2} & X_{t3} & \ldots & X_{tk} \end{pmatrix}$$

Equation (1) may be written as

$$Y_t = x_t \beta + u_t \quad , t = 1, \ldots, T \qquad (1b)$$

Note that equation (1) and (1b) are identical.

Now define a T×1 vector of all T observations on Y_t:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_T \end{bmatrix}$$

Then a T×1 vector of disturbances:

$$u = \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_T \end{bmatrix}$$

And finally X, a T×k matrix of T observations on each of k explanatory variables:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_T \end{bmatrix} = \begin{bmatrix} 1 & X_{12} & X_{13} & . & . & X_{1k} \\ 1 & X_{22} & X_{23} & . & . & X_{2k} \\ . & & . & & & . \\ . & & & . & & . \\ . & & & & . & . \\ 1 & X_{T2} & X_{T3} & . & . & X_{Tk} \end{bmatrix}$$

An individual element of the matrix X can be labeled X_{ij}, where i denotes the row number (= observation number) and j the column number(= variable number). So, for example, X_{32} refers to the third observation [t=3] on the second variable (X_2) in our data set.

Putting this all together we have:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_T \end{bmatrix} = \begin{bmatrix} 1 & X_{12} & X_{13} & . & . & X_{1k} \\ 1 & X_{22} & X_{23} & . & . & X_{2k} \\ . & . & . & & . & . \\ . & . & & . & . & . \\ . & . & & & . & . \\ 1 & X_{T2} & X_{T3} & . & . & X_{Tk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_T \end{bmatrix}$$

which, more compactly, is

$$\mathbf{Y} = \mathbf{X\beta} + \mathbf{u}$$

The classical linear regression model,
$y= f(x_1,x_2,\ldots,x_k, \beta_1, \beta_2,\ldots,\beta_k)+\varepsilon$,
is defined in terms of a number of assumptions about the disturbance term **ε** and also about **X**.

The classical assumptions are:

(a) The functional form of the model is linear. ***Linearity*** means it is *linear in parameters*.

(b) The conditional mean function, E[y | **x**], is different for different set of parameters and for any set of n observations.

(c) The disturbance term **ε** has a zero mean, E[**ε** | for all x observations)=0, ensuring that all data in **X** matrix never provide information about the mean of the disturbance term.

(d) Explanatory variables are uncorrelated with the disturbance term, i.e. Cov[X, **ε]=0.**

(e) The disturbance term has a constant variance and it is not correlated. This establishes the minimum variance property of OLS, such that

$$E\left[\varepsilon_i \varepsilon_j\right] = \sigma^2 I_n \text{ if } i = j, \text{ or } = 0 \text{ if } i \neq j.$$

(f) **X** is a matrix of observations on (k-1) nonstochastic (fixed) variables with **full column rank k** (n>k).

Full column rank ensures the linear independency, and **nonstochastic X** matrix implies that all the explanatory variables remain unchanged. This assumption establishes the result that Var(Y)= $\sigma^2 I_n$ = Var(**ε**).

(g) The disturbance term has a multivariate normal distribution with zero mean and $\sigma^2 I_n$ variance.

Least squares estimation

Least squares residuals are defined by $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$, where $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$.

We can view the vector $\hat{\boldsymbol{\beta}}$ as providing estimators of the elements of the unknown vector $\boldsymbol{\beta}$.

The ***criterion of least squares*** is defined in terms of *minimizing* the residual sum of squares which can be written as $\Sigma e_i^2 = \mathbf{e'e}$.

The residuals sum of squares defined as

$$\mathbf{e'e} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = (\mathbf{y}' - \hat{\boldsymbol{\beta}}'\mathbf{X}')(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

$$= \mathbf{y'y} - \hat{\boldsymbol{\beta}}'\mathbf{X'y} - \mathbf{y'X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\mathbf{X'X}\hat{\boldsymbol{\beta}}$$

Finding the solution vector for $\hat{\boldsymbol{\beta}}$ which minimizes $\mathbf{e'e}$, requires that we differentiate the expression with respect to $\hat{\boldsymbol{\beta}}$ and set the resultant partial derivative equal to a vector of zeros.

$$\frac{\partial(\mathbf{e'e})}{\partial\hat{\boldsymbol{\beta}}} = -2\mathbf{X'y} + 2\mathbf{X'X}\hat{\boldsymbol{\beta}} = \mathbf{0} \Rightarrow (\mathbf{X'X})\hat{\boldsymbol{\beta}} = \mathbf{X'y}.$$

Assumption (f) implies that $\mathbf{X'X}$ is nonsingular and hence we can write for the OLS estimator vector:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y}$$

Two useful results follow from the above formula for $\hat{\boldsymbol{\beta}}$:

Result.1) Using the definition $\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e}$, we can write:

$\hat{\beta} = (X'X)^{-1}X'(X\hat{\beta} + e) = (X'X)^{-1}X'X\hat{\beta} + (X'X)^{-1}X'e$

$= \hat{\beta} + (X'X)^{-1}X'e,$

which yields the result $X'e = 0$.

From this we can deduce that: **(i)** OLS residuals sum to zero and **(ii)** they are uncorrelated with each of the explanatory variables.

Result.2) Given the model $y = X\beta + \varepsilon$, and the formula for the OLS estimator $\hat{\beta} = (X'X)^{-1}X'y$, then we obtain that

$\hat{\beta} = (X'X)^{-1}X'(X\beta + \varepsilon) = \beta + (X'X)^{-1}X'\varepsilon.$

Now consider *the key features of the joint probability distribution for* $\hat{\beta}$.

Unbiasedness and Efficiency

Given the model $y = X\beta + \varepsilon$, and the formula for the OLS estimator $\hat{\beta} = (X'X)^{-1}X'y$, we obtained

$\hat{\beta} = (X'X)^{-1}X'(X\beta + \varepsilon) = \beta + (X'X)^{-1}X'\varepsilon.$

On the assumptions that ε is multivariate normal and that X is nonstochastic, then we can say that $\hat{\beta}$ is also multivariate normal and hence that each estimator $\hat{\beta}_j$ is univariate normal ($j = 1, ..., k$).

Furthermore, since β is a vector of constants and X is non-random, $\hat{\beta}$ is related to both y and ε in a linear way. Each value of $\hat{\beta}_j$ is a linear function of the Y-values and so $\hat{\beta}$ is called a linear estimator of β.

The mean is:

$E(\hat{\beta}) = E(\beta) + E[(X'X)^{-1}X'\varepsilon] = \beta + (X'X)^{-1}X'E(\varepsilon)$, since X is non-random.

Hence, if $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, then $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ which is the definition of an **unbiased estimator**.

In order to consider other properties we need to derive the variance-covariance matrix for $\hat{\boldsymbol{\beta}}$. The variance is:

$$V(\hat{\boldsymbol{\beta}}) = E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})']$$
$$= E[(\mathbf{X'X})^{-1}\mathbf{X'\varepsilon\varepsilon'X}(\mathbf{X'X})^{-1}]$$

since $\mathbf{X'X}$ is symmetric.

$$= (\mathbf{X'X})^{-1}\mathbf{X'}E[\boldsymbol{\varepsilon\varepsilon'}]\mathbf{X}(\mathbf{X'X})^{-1}$$
$$= (\mathbf{X'X})^{-1}\mathbf{X'}.\sigma^2\mathbf{I}.\mathbf{X}(\mathbf{X'X})^{-1}$$
$$= \sigma^2(\mathbf{X'X})^{-1}.$$

The Gauss-Markov theorem proves that, under certain conditions, $\hat{\boldsymbol{\beta}}$ is the best linear unbiased estimator of $\boldsymbol{\beta}$; in this connection "best" means "minimum variance.

When the normality assumption is added, the Cramer-Rao theorem also shows that $\hat{\boldsymbol{\beta}}$ is the efficient estimator.

We can conclude that, under the stated assumptions, the estimator $\hat{\beta}_j \sim N(\beta_j, \sigma^2.a_{hjj})$ where a_{jj} denotes the element in the main diagonal, position j, of $(\mathbf{X'X})^{-1}$.

Hence we can write:

$$\frac{(\hat{\beta}_j - \beta_j)}{\sqrt{\sigma^2\, a_{jj}}} \sim N(0,1).$$

Derivation of an estimator for σ^2:

The variance term σ^2 is unknown and has to be estimated. First, we derive a suitable estimation formula and then, we investigate how this affects the probability distribution derived above.

The OLS residuals are defined as $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ and so we can write:

$$\mathbf{e} = \mathbf{y} - \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'y}$$
$$= [\mathbf{I}_n - \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}]\mathbf{y} = \mathbf{My}$$

where $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}$.

M can be shown to be both symmetric (**M′** = **M**) and *idempotent matrix*, (**M′M** = **M**).

Furthermore,
$\mathbf{MX}=[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{X} = \mathbf{X} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{X}-\mathbf{X} = \mathbf{0}$ (the null matrix).

Premultipling
$\mathbf{y} = \mathbf{X}\beta + \varepsilon$ by **M**
to obtain
$\mathbf{My} = \mathbf{MX}\beta + \mathbf{M}\varepsilon = \mathbf{M}\varepsilon$
since $\mathbf{MX} = \mathbf{0}$.
But $\mathbf{My} = \mathbf{e}$ and so, $\mathbf{e} = \mathbf{M}\varepsilon$.

The sum of squared residuals may be expressed as **e′e,** which therefore equals $\varepsilon'\mathbf{M}'\mathbf{M}\varepsilon = \varepsilon'\mathbf{M}\varepsilon$ since **M** is idempotent.

Now $\mathbf{e}'\mathbf{e} \geq 0$ and hence the quadratic form $\varepsilon'\mathbf{M}\varepsilon \geq 0$. This implies that **M** is a positive semi-definite matrix.

The *trace* (tr) of a (square) matrix is defined as *the sum of the elements contained in the main diagonal*. Special results relate to composite matrices; thus tr(**A**–**B**)=tr(**A**)–tr(**B**), and tr(**ABC**) = tr(**BCA**) so long as **ABC** and **BCA** are permissible.

Consider $\text{tr}(\mathbf{M}) = \text{tr}(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$
$= \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$
$= \text{tr}(\mathbf{I}_n) - \text{tr}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X})$
$= \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{I}_k) = n - k$
Furthermore, the rank of an idempotent matrix can be shown to be equal to its trace, and hence rank(**M**) = n-k.

$E(\mathbf{e}'\mathbf{e}) = E(\varepsilon'\mathbf{M}\varepsilon) = E[\text{tr}(\varepsilon'\mathbf{M}\varepsilon]$, since $\varepsilon'\mathbf{M}\varepsilon$ is a scalar, $= E[\text{tr}(\mathbf{M}\varepsilon\varepsilon')]$.

The operators E and Σ may be taken in either order and trace is simply a form of addition.

$E[\text{tr}(\mathbf{M}\varepsilon\varepsilon')] = \text{tr}[E(\mathbf{M}\varepsilon\varepsilon')] = \text{tr}[\mathbf{M}.E(\varepsilon\varepsilon')] = \text{tr}[\mathbf{M}.\sigma^2\mathbf{I}_n] = \sigma^2.\text{tr}(\mathbf{M}) = \sigma^2(n-k)$.

The **OLS estimator of σ^2** is therefore defined as $\mathbf{s^2 = e'e/(n-k)}$.

This provides the result that $\mathbf{E(s^2) = \sigma^2}$ which means that **s^2 also is an unbiased estimator**.

Finally, it can be shown that

$\mathbf{[(n-k).s^2/\sigma^2] = e'e/\sigma^2 \sim \chi^2(n-k).}$

if s^2 is independent of $\hat{\beta}$.

If we make use of the definition of a "t" random variable here,we can obtain a result that enables us to undertake statistical inferences in relation to **β**.

$$[(\hat{\beta}_j - \beta_j) / \sqrt{(\sigma^2 . a_{jj})}] / \sqrt{[\{(n-k)s^2/\sigma^2\}/(n-k)]}$$

$$= (\hat{\beta}_j - \beta_j) / \sqrt{(s^2 . a_{jj})} \sim t(n-k)$$

The square root of the estimated variance of an estimator is called the standard error (SE).

Tests of Linear Restrictions on Parameters

Consider a linear regression model for testing several **restrictions** on β parameters. The restrictions are tested in:

- a single hypothesis, involving two or more βs such as $\beta_2 = 0$, using **"t-test"**, or
- joint hypotheses, such as $\beta_2 = 0$, $\beta_3 = 0$, **using "F test".**

- **Single Hypothesis: H_0: Rβ=r**

R is a known matrix of order qxk, and r is a column vector. R is assumed to have a full row rank. The number of restrictions denoted by q and parameters by k. **Rβ** is estimated by OLS under the standard CLRM assumptions, and **Rβ** is replaced by $R\hat{\beta}$ in order to test for the validity of restrictions.

The unrestricted model is $\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$, and the OLS estimator is $\hat{\beta} = (\mathbf{X'X})^{-1}\mathbf{X'y}$.

The sampling distribution of $R\hat{\beta}$ is obtained as follows:

$$E(R\hat{\beta}) = R\beta \text{ since R is nonstochastic.}$$

$$Var(R\hat{\beta}) = E[(R\hat{\beta} - R\beta)(R\hat{\beta} - R\beta)']$$

$$= E[R(\hat{\beta} - \beta)(\hat{\beta} - \beta)'R'] = \sigma^2 R (X'X)^{-1}R'.$$

Therefore, $R\hat{\beta} \sim N\{R\beta, \sigma^2 R(X'X)^{-1}R'\}$,
and it is independently of $\{(n-k)s^2/\sigma^2\} \sim \chi^2(n-k)$ so that

$\{R\hat{\beta} - R\beta\}/\sqrt{\{s^2R(X'X)^{-1}R'\}} = \{R\hat{\beta} - R\beta\}/SE(R\hat{\beta}) \sim t(n-k)$.

This result provides the basis for inferences about **Rβ**.

For example, if a restriction is imposed on β_2, such that $\beta_2 = 0$, when k=3, then r=0 and R=[0 1 0].

- **Joint Hypothesis: H_0: Rβ=r**

Consider an **unrestricted model**

$Y_i = \beta_1 + \beta_2X_{2i} + ... + \beta_hX_{hi} + \beta_{h+1}X_{h+1,i} + ... + \beta_kX_{ki} + \varepsilon_i$

and suppose that we wish to test

H_0: $\beta_{h+1} = 0, ..., \beta_k = 0$ against H_1: H_0 untrue, with k-h restrictions. The null hypothesis states that a set of X variables does not affect Y.

This test is conducted by running two regressions, relating to an unrestricted model (UM) and a restricted model (RM).

The **restricted model** is

$Y_i = \beta_1 + \beta_2X_{2i} + ... + \beta_hX_{hi} + \varepsilon_i$

Restrictions are imposed on the UM in order to test the validity of RM. **OLS is used to estimate** both equations and to obtain RSS_u and RSS_r respectively.

If H_0 is true, $X_{h+1}, ..., X_k$ variables are irrelevant variables for the model.
If H_0 is wrong, $X_{h+1}, ..., X_k$ variables are relevant variables for the model and adding them to the model results in an increase in the explanatory power and thus a decrease in RSS.

The *test statistic* is given by

$[(1/\sigma^2)\{RSS_r - RSS_u\}/(k-h)] / [[(1/\sigma^2)\{RSS_u\}/(n-k)] \sim F(k-h, n-k)$ if H_0 true.

For example, consider a unrestricted model $Y_i = \beta_1 + \beta_2X_{2i} + \beta_3X_{3i} + \beta_4X_{4i} + \beta_5X_{5i} + \varepsilon_i$ and test H_0: $\beta_3 = \beta_2$, $\beta_5 = 1 - \beta_4$.
The restricted model is $Y_i = \beta_1 + \beta_2X_{2i} + \beta_2X_{3i} + \beta_4X_{4i} + (1 - \beta_4)X_{5i} + \varepsilon_i$, which may be written as: $(Y - X_5)_i = \beta_1 + \beta_2(X_2 + X_3)_i + \beta_4(X_4 - X_5)_i + \varepsilon_i$

Chow's Parameter Stability Test

One important application of the F test is for **parameter stability**. Suppose that we wish to consider ***whether or not one or more of the parameters has changed in value at some point in the data "time period".***

We can write the unrestricted model as:

$Y_i = \beta_1 + \beta_2 X_{2i} + .. + \beta_k X_{ki} + \varepsilon_i \qquad i = 1,., n_1 \qquad (i)$

$Y_i = \gamma_1 + \gamma_2 X_{2i} + .. + \gamma_k X_{ki} + \varepsilon_i \qquad i = n_1+1, ..., n_1+n_2 \qquad (ii)$

where $n_1 + n_2 = n$.

We wish to test H_0: $\beta_1 = \gamma_1$, $\beta_2 = \gamma_2$, ..., $\beta_k = \gamma_k$ (i.e. k restrictions).

The restricted model is:

$Y_i = \beta_1 + \beta_2 X_{2i} + ... + \beta_k X_{ki} + \varepsilon_i \qquad i = 1, ..., n$

RSS_u is made up of two parts, $RSS_{(i)} + RSS_{(ii)}$, and the degrees of freedom associated with the unrestricted model is $(n_1 - k) + (n_2 - k) = n - 2k$.

Hence *the test statistic* may be written as:

$[(RSS_r - RSS_u)/(k)]/ (RSS_u)/(n-2k)] \sim F(k, n-2k)$ under H_0.

Rejecting H_0 implies that the evidence is consistent with some kind of structural break. *The choice of break point n_1 may either be determined exogenously* (there is some reason to expect a break at this point) *or arbitrarily* (as a routine diagnostic test).

2.2 Inference Under Different Assumptions:

The k variable regression model is

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + \varepsilon_t, \quad t = 1,\ldots,T$$

Assumptions:

(1)The T×k matrix of regressors, X, is non-stochastic, and the regression model is correctly specified.

(2)The X matrix is non-singular (of full rank) so there is no perfect collinearity among the regressors.

(3) E(u) = 0.

(4)Var(u) = $\sigma^2 I$, where I is a T×T identity matrix.

(5)Cov(u_t,u_s)=0 for all s≠t.

(6)The disturbance term u is multivariate normally distributed.

1) The classical linear regression model with non-stochastic regressor variables and normally distributed disturbances.

This case establishes that:

- $\hat{\beta}$ is the minimum variance unbiased estimator of β
- $\hat{\beta}$ is (exactly) normally distributed, with $\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$
- $\hat{\beta}$ is the minimum variance unbiased estimator of β
- The OLS estimator of the disturbance term variance σ^2 is unbiased
- t and F test statistics are valid.

2) The classical linear regression model with non-stochastic regressor variables and non-normally distributed disturbances.

The main difference between (1) and (2) is that:

- The normality of the estimator $\hat{\beta}$ no longer holds in finite samples.
- As t and F test statistics are derived on the assumption that $\hat{\beta}$ is normally distributed, inference based on t and F tests will now only be **approximate** in finite samples.
- Need for ***asymptotic theory***.

3) The classical linear regression model with stochastic and stationary regressors:
If a variable is stochastic, then it is a random variable and thus it has a probability distribution. It is non-stochastic if it not a random variable.

Some variables are **non-stochastic**, including ***intercept, quarterly dummies, dummies for special events and time trends***. In any period, each takes one value known with certainty. However, many economic variables are stochastic

Stochastic Regressors can be viewed in the following two cases:

➔ Consider a regression model with a ***lagged dependent*** variable (Y_{t-1}) as a regressor:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \beta_4 Y_{t-1} + u_t \quad , t = 1,\ldots, T$$

Since $Y_t=f(u_t)$, and thus it is a random variable, and also $Y_{t-1}=f(u_{t-1})$, and so Y_{t-1} is a random or stochastic variable.

➔ Consider a variable measured by a process in which ***random error measurement occurs***.

If a variable is determined by a data generating process and includes a ***chance*** component, then the variable will be stochastic.

Therefore, ***stochastic regressors will be correlated with, or not independent of, the random disturbance***. This possibility does not arise when regressors are non-random.

In addition, consider a sequence of observations on a time series variable, X_t, t = 1, 2, and assume that all regressors are ***stationary***.

The variable X is said to be **weakly or covariance stationary** if all of the following conditions are satisfied:

(i) $E(X_t) = \mu$, with μ a constant finite number;
(ii) $Var(X_t) = \sigma^2$, with σ a finite, positive number;
(iii) $Cov(X_t, X_{t-k}) = \sigma_k$, a constant, finite number, for $k \neq 0$ and for any t.

4) The classical linear regression model with stationary, stochastic regressor variables, and independently identically distributed normal disturbances, independent of the regressors:

Results:

- The OLS estimator is unbiased. This is also valid for the OLS estimators of σ^2 and the OLS standard errors.
- The OLS estimators are efficient.
- The t and F test statistics valid although the regressors are stochastic.

However, the assumption that *the regressors are independent of the disturbances for any size of sample* is very strong, and unlikely to be satisfied in practice.

5) The classical linear regression model with stationary, stochastic regressor variables, and independent identically distributed but non-normal disturbances, independent of regressors:

Results:

- $\hat{\beta}$ is still unbiased even if the disturbances are non-normal.
- The finite sample distributions of the OLS estimator of σ^2 (of s^2) and of the t and F test statistics are no longer standard.
- Need for asymptotic theory in hypothesis testing.
- The OLS estimators are consistent.

A BRIEF VIEW ON THE ASYMPTOTIC THEORY: CONSISTENCY AND CONVERGENCE IN PROBABILITY

Consider a sequence of random variables that depends in some way on the sample size, T.

For example, the sample mean of a series of variables $Y_1, Y_2, ..., Y_T$, defined as

$$\overline{Y}_T = \frac{\sum_{t=1}^{t=T} Y_t}{T}$$ can be thought of as a random sequence. We can imagine a sequence of the random variables $\overline{Y}_T$ for various sample sizes: $\overline{Y}_1\ \overline{Y}_2\ \ldots\ \overline{Y}_{40}\ \ldots\ \overline{Y}_{110}$

Convergence in probability **:**

Let X_T represent a sequence of random variables. The random sequence X_T converges in probability to a constant X if, as $T \to \infty$

$$\underset{T\to\infty}{limit} \Pr\left(|X_T - X| > \partial\right) = 0 \text{ for any } \partial > 0$$, where the symbol | | denotes the absolute value.

This idea can also be expressed in two other ways:

- $X_T \xrightarrow{P} X$ where the arrow with P above it reads as "converges in probability".
- plim(X_T) = X

A consistent estimator

Now consider $\hat{\beta}$, an estimator of β. An estimator is said to be consistent ***if it converges in probability to the true parameter value***. Thus we can say that:

$\hat{\beta}$ is a consistent estimator of β if

$$\underset{T\to\infty}{limit} Pr\left(|\hat{\beta} - \beta| > \partial\right) = 0 \text{ for any } \partial > 0$$

Alternatively, we could write that $\hat{\beta}$ is a consistent estimator of β if

$$\hat{\beta} \xrightarrow{P} \beta \quad \text{OR} \quad \text{plim}(\hat{\beta}) = \beta$$

Consistency of the OLS estimator of β.

$$\hat{\beta} = \beta + (X'X)^{-1}X'u$$

Since
$$= \beta + \left(\frac{1}{T}(X'X)\right)^{-1}\frac{1}{T}X'u$$

Taking the probability limit of this expression:

$$plim(\hat{\beta}) = \beta + plim\left[\left(\frac{1}{T}\right)^{-1}\right].plim\left[\frac{1}{T}\right]$$

$$plim(\hat{\beta}) = \beta + plim\left[\left(\frac{1}{T}\right)^{-1}\right].plim\left[\frac{1}{T}\right]$$

If we assume that $plim\left(\frac{1}{T}X'u\right) = 0$, then $plim\left(\frac{1}{T}\right) = 0$. Hence the OLS estimator is consistent for β.

It is important to note that our proof has assumed that $plim\frac{X'X}{T}$ exists and is a non-singular matrix. This requires that the variables in X be **stationary**.

The intuition is that as the sample size grows to infinity, the probability of $\hat{\beta}$ differing from β in absolute value by more than any positive amount, no matter how small, collapses to zero. Note, however, that if an estimator is consistent, that does not tell us anything about whether or not the estimator is **biased** in a finite sized sample, nor anything about how large that bias may be.

It is also worth noting that an estimator $\hat{\beta}$ will be consistent if, as $T \to \infty$, $E(\hat{\beta}) \to \beta$, and $Var(\hat{\beta}) \to 0$

However, although this pair of conditions is sufficient for consistency, they are not necessary for it. That is, together they guarantee consistency; but if they are not both present, it is still possible that an estimator is consistent.

Stationarity assumption implies that $\frac{1}{T}\sum X_T X_T' \xrightarrow{P} Q$ where Q is a positive definite matrix.

Given this, it can be established that $\sqrt{T}\left(\hat{\beta}_T - \beta\right) \xrightarrow{D} N\left(0, \sigma^2 Q^{-1}\right)$ where the arrow symbol with a D above it is read as "converges in distribution to". Put another way, what this says is that as the sample size grows ever larger, then in the limit the term on the left hand side becomes arbitrarily close to a random variable with distribution as given on the right-hand side of the expression.

Alternatively, this could be written as $\hat{\beta} \sim^a N(\beta, \sigma^2(X/X)^{-1})$. Given this, a t test statistic will be asymptotically distributed as N(0,1) , an F test statistic for m restrictions will be asymptotically distributed as F(m, T-k).

6) The linear regression model with stationary, stochastic regressor variables and iid disturbances, X and u are asymptotically uncorrelated:

The assumption that ***the equation disturbances and the regressors are independent*** is an extremely strong assumption.

For example, suppose that the regressors are independent of the contemporaneous disturbance but, not of preceding disturbances.

In this case, however, the equation disturbances and the regressors are uncorrelated in the limit as the sample size goes to infinity, even though they are correlated in finite size samples.

Assume that the equation disturbances and the regressors are uncorrelated in the limit as the sample size goes to infinite, then it is said that "**the regressors are asymptotically uncorrelated with the disturbances**".

Formally, we have: $$plim\left(\frac{1}{T}\right)=0$$

or in matrix terms $$plim\left(\frac{1}{T}X'u\right)=0$$

Results:

- The OLS estimator will be biased in finite samples.
- The OLS estimator has some desirable 'large sample' or asymptotic properties, being consistent and asymptotically efficient.
- The OLS estimators of the disturbance variance and so of coefficient standard errors will also have desirable large sample properties.

7) The linear regression model with stationary stochastic regressor variables, iid disturbances, and X and u are correlated even asymptotically:

If $\text{plim}\left(\frac{1}{T}X'u\right) \neq 0$, then $\text{plim}\left(\hat{\beta}\right) \neq \beta$, and hence OLS is inconsistent.

There are three possible causes of correlation of regressors and disturbances:

(1) Simultaneous equations bias
(2) Errors in variables

(3) The model includes a lagged dependent variable and has a serially correlated disturbance.

Suppose that the following regression model is estimated by OLS
$Y = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + u_t$, in which $u_t = \rho u_{t-1} + \varepsilon_t$

with ρ non-zero. By lagging the equation for Y by one period, it is clear that Y_{t-1} is correlated with u_t irrespective of the sample size (given that $\rho \neq 0$).

8) The linear regression model with non-stationary stochastic regressors:

The regression model will contain stochastically non-stationary regressors if one or more of the following conditions is not satisfied:

(i) $E(X_t) = \mu$, with μ a constant finite number;
(ii) $Var(X_t) = \sigma^2$, with σ a finite, positive number;
(iii) $Cov(X_t, X_{t-k}) = \sigma_k$, a constant, finite number, for $k \neq 0$ and for any t.

2.3 Model Specification and Misspecification Testing

The estimates derived from linear regression techniques and inferences based on those estimates, are only valid under certain assumptions. If these certain conditions are satisfied, then the empirical model will be a **"well-specified" model**.

In order to construct a well-specified empirical model, **an appropriate procedure for estimation and testing includes 5 steps as follows**:

(1) Specify a statistical model that is consistent with the relevant prior theory, in the sense that it embodies the theoretical relationship that the researcher believes exists between a set of variables. Notice that this first step requires that at least two choices be made:

(i) The choice of the **set of variables** to include in the model.

(ii) The **choice of functional form** of the relationship.

(2) Select an estimator possessing certain desired properties provided the regression model satisfies a particular set of assumptions. In many circumstances, the estimator selected will be the OLS estimator. The OLS estimator is known to be BLUE under the validity of a particular set of assumptions.

Even under less restrictive assumptions, the OLS estimator **may** still be the most appropriate one to use.

However, there may be circumstances where we shall wish **to use some other estimator**. We shall denote the regression model as a ***statistically well-specified*** one for a given estimator if all of the assumptions are satisfied.

The regression model will be called **statistically misspecified** if one or more of the assumptions are not satisfied.

(3) Estimate the regression model using the chosen estimator.

(4) Test whether the assumptions made are valid (in which case the regression model is statistically well-specified) and the estimator will have the desired properties.

(5) (a) If these tests show no evidence of misspecification in any relevant form, go on to conduct statistical inference about the parameters.

OR

(b) If these tests show evidence of misspecification in one or more relevant forms, then it **may** be possible to find an alternative estimator.

Testing The Assumptions of The Linear Regression Model

The process of testing the model assumptions is known as **misspecification (or diagnostic) testing**. The misspecification tests are for assumptions about:

I. The Specification of the Regression Model
II. The Equation Disturbance Term
III. The Parameters of The Model
IV. The Asymptotic Correlation Between Regressors and Disturbance Terms.
V. The Assumption of Stationary Regressors.

I. The Specification of the Regression Model

(A) The Choice of Variables:

- Omission of a relevant variable from the model.
- Inclusion of an irrelevant variable in the model.

		TRUE	MODEL
		$Y_t = \beta X_t + u_t$	$Y_t = \beta X_t + \gamma Z_t + u_t$
ESTIMATED	$Y_t = \beta X_t + u_t$	**A**	**B**
REGRESSION	$Y_t = \beta X_t + \gamma Z_t$	**C**	**D**

Results:

Case A: True model estimated

⇒ β and σ^2 estimated without bias and efficiently.

⇒ $SE(\hat{\beta})$ is correct standard error, and so use of t and F tests is valid.

Case D: True model estimated

⇒ β, γ and σ^2 estimated without bias and efficiently.

⇒ Standard errors are correct, so use of t and F tests valid.

Case B: Wrong model estimated due to omission of relevant variable.

⇒ $\hat{\beta}$ is biased. [In the special case where X and Z are uncorrelated in the sample, $\hat{\beta}$ is unbiased].

⇒ $SE(\hat{\beta})$ is biased, as is the OLS estimator of σ^2

⇒ Use of t and F tests are not valid.

Case C: Wrong model estimated due to inclusion of an irrelevant variable.

$\Rightarrow \hat{\beta}$ is unbiased but inefficient (relative to the OLS estimator that arises when the true restriction is imposed (as in case A).

$\Rightarrow SE(\hat{\beta})$ is biased, as is the OLS estimator of σ^2

$\Rightarrow$ Use of t and F tests are not valid.

(B) The Choice of Functional Forms:

Assume that

the estimated model is $Y_t = \beta_1 + \beta_2 X_t + u_t$,

but the true model is $Y_t = \beta_1 + \beta_2 ln(X_t) + u_t$.

Thus the *functional form* of the model is not correctly specified.

The **consequence** of this will be estimating the wrong model. The inference will be misleading and also meaningless.

Suppose Y=f(X) and **the correct functional form** could be in a form of

$Y = \beta_1 + \beta_2 X$	**LINEAR**
$\ln(Y) = \alpha_1 + \alpha_2 \ln(X)$	**LOGARITHMIC (LOG-LINEAR)**
$Y = \gamma_1 + \gamma_2 \ln(X)$	**LIN-LOG (SEMI-LOG)**
$\ln(Y) = \delta_1 + \delta_2 X$	**LOG-LIN (SEMI-LOG)**
$Y = \varphi_1 + \varphi_2 \frac{1}{X}$	**RECIPROCAL/RATIO FORM**
$\frac{1}{Y} = \theta_1 + \theta_2 X$	**RECIPROCAL/RATIO FORM**
OR $Y = \frac{1}{\theta_1 + \theta_2 X}$	
$Y = \psi_1 + \psi_2 X + \psi_3 X^2 + \ldots$	**POLYNOMIAL**

Each of these forms implies a different type of relationship between Y and X.

For example, if the log-linear model in fact correctly describes the Y, X relationship, then there is a constant elasticity of Y with respect to X, given by α_2.

In other terms, the estimates will be expressed in *percentages* in a log-linear model. In contrast, the linear form implies that there is not a constant elasticity of Y with respect to X. Then, the estimates will be expressed in terms of *unit*.

One way of investigating the appropriateness of our choice of functional form is by using Ramsey's RESET test.

II. The Equation Disturbance Term

A. Serially Correlated Disturbance Term

Assume that the specification of the regression model with serially correlated error term is

$$Y_t = \beta_1 + \beta_2 X_{2t} + ... + \beta_k X_{kt} + u_t \text{ and } t = 1,..., T$$

Possible **consequences** are:
⇒ The OLS estimators are not BLUE.
⇒ The OLS estimator is inefficient although still unbiased. *Stating that an estimator is inefficient means that the estimator has a higher variance than some other unbiased estimator.* If the error term did in fact exhibit serial correlation, the OLS estimator would be making no use of this information.
⇒ The standard errors of the OLS estimators are biased. This means that the use of t and F statistics to test hypotheses is misleading or invalid.

The **Generalised Least Squares (GLS) estimator**- could be used instead of OLS, and would yield unbiased and efficient parameter estimates.

If the regressors are stochastic, serial correlation will also result in efficiency losses, and may also lead to OLS being inconsistent.

Consider the following specification of the equation disturbance term:

$$u_t = \rho u_{t-1} + \varepsilon_t$$

where ε_t is assumed to be a **"white noise" error** term (that is, it has a zero mean, is serially uncorrelated and has a constant variance).

This equation states that the disturbance term u_t is generated by a **first order serially correlated (autoregressive) process**. That is, **an AR(1) process**, where *AR* is used to mean *autoregressive*. If $\rho=0$, there is no serial correlation, so that $u_t = \varepsilon_\tau$.

Detecting Serial Correlation:

- Visual inspection of a graph of disturbances.
- The **Durbin-Watson** test can be used to test the null hypothesis that the error term is serially uncorrelated, against the alternative that each disturbance term is correlated with the disturbance term in the previous period. In this alternative case, the disturbance is serially correlated of order one.

The required null and alternative hypotheses for this test are:

$$H_0 : \rho = 0$$

$$H_A : \rho \neq 0$$

The **DW statistic** is defined as $$DW = \frac{\sum_{t=2}^{t=T} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{t=T} \hat{u}_t^2}$$

Note that DW is only a test for **first order autoregression**. In addition, the DW statistic is only valid in a model in which no lagged values of the dependent variable are included as regressors.

- **The Durbin h test** is a test for an AR(1) error process in a model including lagged dependent variables. The **statistic** is defined as

$$h = (1 - DW/2)\sqrt{T / [1 - T\, V(\hat{\lambda})]}$$

where $V(\hat{\lambda})$ is the estimated variance of the OLS estimator of the coefficient on the lagged dependent variable. The Durbin h statistic is distributed asymptotically as a standard normal random variable if the null of no serial correlation is correct.

- **Godfrey's Lagrange Multiplier (LM) Tests** is for a more general form of process for the equation disturbance term, in which the disturbance term is a p^{th} order autoregressive process (AR(p)). That is

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_p u_{t-p} + \varepsilon_t$$

B. Non-Constancy of the Disturbance Variance

The classical linear regression model assumes that the variance of the equation disturbance term is constant over the whole sample period. That is

$$\sigma_t^2 = \sigma^2 \quad for\ all\ t,\ t = 1,...,T$$

If this assumption is false, then

⇒ OLS estimator is no longer efficient.
⇒ The use of t and F statistics to test hypotheses is misleading or invalid.

If the researcher knew the true structure of the **heteroscedasticity**, then the Generalised Least Squares (GLS) estimator could be used instead of OLS, and would one again yield unbiased and efficient parameter estimates.

If the assumption of **homoscedasticity** is false, then by definition the disturbance terms are heteroscedastic.

There are many **reasons** that the disturbance term could be heteroscedastic.

- The variance might be **related to the value taken by one variable**, Z, which may or may not be included as a variable in the regression model.
- The variance of the equation disturbance term is linearly related to the values taken by a set of p variables, an intercept plus p-1 other variables denoted Z_2 to Z_p.

An **F test** can then be used to test the validity of the p-1 restrictions stated under the null hypothesis. This test will only be an approximate one in finite samples.

In particular, the disturbance ε cannot be normal if the disturbance of the original regression, u, is normal. The test can also be conducted using an alternative statistic.

If R^{*2} denotes the R^2 statistic from the second stage (auxiliary) regression, then it is approximately true in finite samples that

$$T \bullet R^{*2} \sim \chi^2_{p-1}$$

if the null hypothesis is valid. It follows that the null of homoscedasticity would be rejected if the value of TR^{*2} exceeded the critical value of the chi-square distribution with p-1 degrees of freedom at the chosen significance level.

However, if you had no prior idea about which variables could influence the variance, then the disturbance process can be represented by the following expression:

$$Var(u_t) = E(u_t^2) = \sigma_t^2 = \sigma^2 + \gamma(E[Y_t])^2$$

A test could be implemented by:
(1) Estimating the **original** regression by OLS,
(2) Saving the fitted values and the estimated residuals, then
(3) Using the residuals as proxies for the unobserved disturbances and the squares of the fitted values as proxies for $[E(Y)]^2$, and finally **(4)** Estimate the following auxiliary regression:

$$\hat{u}_t^2 = a + \gamma \hat{Y}_t^2 + \varepsilon_t \ , \ t = 1,..,T$$

For the hypotheses H_0: $\gamma = 0$ against H_1: $\gamma \neq 0$

The statistic TR^2 (where R^2 is from the **auxilliary** regression) has, under the null hypothesis (γ=0) a chi-square distribution with one degree of freedom. This is known as the **LM (Lagrange multiplier)** version of the test. An even simpler alternative would be to use an **F test (or t test)** statistic for the null hypothesis that γ=0. In this case, the F statistic would have an F distribution with 1 and T-k degrees of freedom if the null hypothesis is correct.

The **Breusch Pagan test** and **Goldfeld Quandt** test can also be used for testing heteroscedasticity.

Engle (1982) suggested that many time series exhibit a special form of heteroscedasticity.
This is **autoregressive conditional heteroscedasticity**.

C. Non-Normality of Disturbance Term

If the equation disturbance terms are not normally distributed, then it is not possible to derive exact distributions for the estimators and other related statistics for finite size samples.

The assumption that the equation disturbances are normally distributed can be tested using a procedure suggested by **Jarque and Bera** (1981).

Any probability distribution can be characterized by values taken by its moments. The **first moment** is its mean or expected value, the **second moment** its variance. The **third moment** and **fourth moment** are sometimes described as the degree of skewness and kurtosis of the distribution.

III. The Parameters of The Model:

It is also important to note that in estimating, for example, the linear regression model Y = Xβ + u, and we are not only assuming constancy of the parameter set of β, but we are also assuming that *the variance of the disturbance term, σ^2, is a constant number over the whole sample*.

The assumption of ***parameter stability*** is tested by the **Chow Parameter Constancy test**.

It is implemented by dividing the sample into two sub-samples, estimating each sample separately by OLS, and then testing whether the two sets of parameter estimates are significantly different from one another. This can be tested using a conventional F test procedure. The Chow test is a commonly used parameter stability test.

In addition, the **Chow Predictive Failure test** checks parameter constancy with in-sample prediction. Suppose that we divide the full sample of T observations into two subsamples, T-m observations called the estimation period and m observations called the forecast period.

The test statistic is obtained as the F test statistic. If the F statistic exceeds the relevant critical value, we reject the null of parameter constancy.

2.4 Non-Spherical Disturbances

The classical regression model we have considered so far maintains that the process generating the observations on the dependent variable Y is adequately represented by

$\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$

where a number of assumptions are made concerning the regressors and the random variable ε. On the basis of these assumptions, the disturbance term can be specified as $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_v)$.

However, the disturbance variance becomes $\mathbf{V(\varepsilon) = \sigma^2\Omega}$, where **Ω** is positive definite and σ^2 is an unknown parameter, rather than $\sigma^2\mathbf{I_n}$ for <u>two reasons</u>:

(1) Heteroscedastic errors: $var(\varepsilon_i) \neq var(\varepsilon_j)$ for at least one i and j; i.e. the disturbance term variance is not constant across all observations.
(2) Cov $(\varepsilon_i, \varepsilon_j) \neq 0$, $i \neq j$, for at least one i and j; i.e. the disturbances are autocorrelated or serially correlated (correlated across observations).

<u>Result:</u> $\hat{\beta}$ is remains unbiased estimator of β, but inefficient. To prove this is

$$\hat{\beta} = (\mathbf{X'X})^{-1}\mathbf{X'y}$$
$$= (\mathbf{X'X})^{-1}\mathbf{X'}(\mathbf{X\beta} + \boldsymbol{\varepsilon})$$
$$= \boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'\varepsilon} \quad \text{and hence}$$

$$E(\hat{\beta}) = \boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'}E(\boldsymbol{\varepsilon}) \quad \text{since } \mathbf{X} \text{ is non-random}$$
$$= \boldsymbol{\beta} \quad \text{since } E(\boldsymbol{\varepsilon}) = \mathbf{0}\ .$$

$$V(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']$$
$$= E[(\mathbf{X'X})^{-1}\mathbf{X'\varepsilon\varepsilon'X}(\mathbf{X'X})^{-1}]$$
$$= (\mathbf{X'X})^{-1}\mathbf{X'}.E(\boldsymbol{\varepsilon\varepsilon'}).\mathbf{X}(\mathbf{X'X})^{-1}$$
$$= (\mathbf{X'X})^{-1}\mathbf{X'}.(\sigma^2\boldsymbol{\Omega}).\mathbf{X}(\mathbf{X'X})^{-1}$$
$$= (\mathbf{X'X})^{-1}\mathbf{X'VX}(\mathbf{X'X})^{-1}.$$

Note that the "conventional" estimator of the covariance matrix of $\hat{\beta}$ is $\sigma^2(\mathbf{X'X})^{-1}$, which is incorrect in present circumstances. Hence the conventional estimator for $V(\hat{\beta})$ is invalid.

There is another estimator of **β**, called the **generalised least squares** (GLS) estimator, denoted by $\tilde{\beta}$, which is also unbiased but has a lower variance than $\hat{\beta}$. The GLS estimator is given by

$$\tilde{\beta} = (\mathbf{X'V^{-1}X})^{-1}\mathbf{X'V^{-1}y} = (\mathbf{X'\Omega^{-1}X})^{-1}\mathbf{X'\Omega^{-1}y}$$

This estimator is unbiased and fully efficient, but can only be used if **Ω. is known.**

If **Ω is unknown**, we can proceed only by replacing it by an estimate, say **Ω***. The resulting estimator $\tilde{\beta}^*$ is known as the **feasible generalised least squares (FGLS)** estimator. If **Ω*** is a consistent estimator of **Ω**, it can be shown that $\tilde{\beta}^*$ is a consistent estimator of **β**.

Derivation of the GLS estimator

Given the definition $V(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{\Omega}$ positive definite matrix, a nonsingular matrix **P** of order n can be found such that $\mathbf{P'P} = \boldsymbol{\Omega}^{-1}$ and $\mathbf{P\Omega P'} = \mathbf{I_n}$.

Premultiplying the regression model by **P**, **the transformed model** is obtained as

$$\mathbf{Py} = \mathbf{PX}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\varepsilon}$$

so that

$E(\mathbf{P}\boldsymbol{\varepsilon}) = \mathbf{P}.E(\boldsymbol{\varepsilon})$ and

$V(\mathbf{P}\boldsymbol{\varepsilon}) = \mathbf{P}(\sigma^2\boldsymbol{\Omega})\mathbf{P'} = \sigma^2\mathbf{P\Omega P'} = \sigma^2\mathbf{I_n}$.

So the covariance matrix of the transformed model satisfies classical assumptions, implying that the OLS estimators of the parameters of the transformed model will be unbiased and efficient.

Applying OLS to the transformed model, to obtain the GLS estimator:

$$\tilde{\beta} = [(\mathbf{PX})'(\mathbf{PX})]^{-1}(\mathbf{PX})'(\mathbf{Py}) = [\mathbf{X'P'PX}]^{-1}\mathbf{X'P'Py}$$
$$= (\mathbf{X'\Omega^{-1}X})^{-1}\mathbf{X'\Omega^{-1}y}$$

The variance-covariance matrix of $\tilde{\beta}$ is given by:

$$V(\tilde{\beta}) = \sigma^2[(\mathbf{PX})'(\mathbf{PX})]^{-1} = \sigma^2(\mathbf{X'\Omega^{-1}X})^{-1} = (\mathbf{X'V^{-1}X})^{-1}$$

and since $\tilde{\beta}$ is efficient this expression provides the minimum variances for estimators of β.

The GLS estimator of σ^2 is unbiased, since

$\tilde{\sigma}^2 = [(\mathbf{Py} - \mathbf{PX}\tilde{\beta})'(\mathbf{Py} - \mathbf{PX}\tilde{\beta})]/(n-k)$ where

$E(\tilde{\sigma}^2) = \sigma^2$.

An alternative expression, written in terms of the transformed model, is given by

$[(\mathbf{y} - \mathbf{X}\tilde{\beta})'\mathbf{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\beta})] / (n-k)$.

Note again that the OLS estimator of σ^2, s^2, will not be an unbiased estimator in the situation where $V(\boldsymbol{\varepsilon}) \neq \sigma^2\mathbf{I_n}$.

3. Econometric Modelling:

3.1 Model Selection:

A primary concern in econometric modelling is to find a suitable model for a given time series. Model selection depends mainly on ***the properties of the time series, the number of observations, and the reason for evaluating the model. In*** general **there are three stages in model construction.** These are

(1) Model specification
(2) Model estimation
(3) Model evaluation

Box and Jenkins (1970) used the ARIMA models for sample sizes being more than fifty. The major contribution of Box and Jenkins has been to provide a general strategy for model building. The Box and Jenkins model selection approach has been widely used in the estimation and testing process in econometric studies. Box and Jenkins suggest a **three-step approach to pure time series modelling**:

1. ***Identification*:** A tentative autoregressive integrated moving average (ARIMA) model is specified through examination of the correlogram and the partial autocorrelation functions.

A *correlogram* shows no sign of decrease in the absolute magnitude of the estimated autocorrelation coefficients if the series are non-stationary.
If the estimated autocorrelation coefficients do not show a sign of decay, then the data must be transformed to ensure stationarity.

The **stationarity transformation** with economic time series is first, taking logarithms and then, differencing. Thus, if the data generating process for a time series is non-stationary, Box and Jenkins suggest differencing the data until they are stationary.

Once stationarity is achieved, the next step is to identify the order of the estimation process.

- For a **pure moving average process of order q, MA(q),** the correlogram will display the estimated autocorrelation coefficients which are significantly different from zero only up to lag q, whereas the partial autocorrelation function will diminish slowly in size.

- For a pure **autoregressive process of order p, AR(p)**, the correlogram will reveal that the estimating autocorrelation coefficients diminish slowly, as the estimated autocorrelation coefficients will tend to be significantly zero beyond lag p.

- For a mixed process neither the correlogram nor the partial correlation coefficient reveals a cut off.

2. ***Estimation*:** Estimating the parameters of a pure AR(p) model by least squares and minimizing the residual sum of squares, or by the maximum likelihood methods. Non-linear techniques must be used for MA(q) and mixed processes.

3. ***Diagnostic checking*:** Diagnostic checking provides the evidence about the randomness of the fitted residuals. If the estimated parameters of the estimated model are significantly different from zero and the fitted residuals appear to be white noise, the ARIMA model should be used as an adequate model

Hendry and Richard (1983) suggest six criteria for model selection.

(i) Models must be *data admissible.*
(ii) Models must be *consistent with theory.*
(iii) In a single equation model, regressors should be at least *weakly exogenous*. If the regressors are not weakly exogenous, then they must be treated as endogenous variables and they must be modelled in a system equation context. In a simultaneous equation model, regressors should be in the reduced form.
(iv) Models should have constant parameters. *Parameter constancy* in the data generating process is related to constancy of parameters over the sample period and to invariance regarding an intervention. Parameter constancy is a fundamental requirement for econometric modelling and its implications. It is crucial for prediction, testing and analysing economic policy.
(v) A model must be *data coherent.* This requires that the difference between the actual and fitted values should be random. In other words, in a data coherent model, the error process should be white noise.
(vi) A satisfactory model should *encompass* rival models. That is, an empirical model must be able to explain the relevant characteristics of other models.

Hendry and Wallis (1984) contribute to the econometric modelling approach in several aspects. It is very important to distinguish among the various kinds of information available for econometric studies in order to achieve the correct formulation of an econometric model.

In econometric modelling the available information is obtained from
- The data sources that are partitioned relative to any point in time t into the past, the present, and the future.
- Theoretical and other prior information.
- The structure of the economic system.
- The rival models.

It is noted that the appropriate criteria to each class of information have been developed with respect to the following concepts:

- Models must be designed to satisfy the **absence of residual serial correlation and residual heteroscedasticity** by appropriate specification of the lag structure and the functional form.

Thus, the fitted model should not deviate from the observed data systematically and heterogeneously to describe the relative past.

- Explanatory variables in an econometric model should be **exogenous** in order to allow the use of contemporaneous effects.

- Either the parameters of interest or the goodness of fit statistics should be **time invariant** to ensure the adequate characterization of future data, and to avoid the predictive failure in the model.

- The estimated coefficients should be tested directly or indirectly to achieve the **theory consistency.**

- **Measurement information** can be carried out directly, or imposed by corrections, logarithms, etc.

- The estimated model could **encompass the results of rival models** to obtain the sufficient model. This is a necessary, but not sufficient condition for an adequate model.

Hendry (1995) presents the priority of empirical econometric modelling in the systematic and integrated framework. Since the empirical models are constructed by using theories and observable data sets, the main problem is to find an empirically relevant model for understanding economic behaviour.

Hendry explains the main roles of econometric models in economics as follows:

1) Econometric models are the data summaries.
2) They provide economic interpretation of empirical evidence.
3) They play an important role in evaluating the relative explanatory power of several competing economic theories.
4) They are the main source for understanding how economies work.

Time series data exhibits a wide variety of patterns, such as **trend, seasonality, and volatility**. In modelling these patterns, there are many competing econometric models and modelling techniques.

The question is whether econometric models should be constructed on pure economic theory or on statistical properties?

It is obvious that neither could explain the real economic activities. Thus the bridge between these approaches is to find an **"adequate" model**.

In order to find an adequate model, it is important to state clearly the difference between the "theory models" and the "empirical models".

Theory models postulate theory relationship between latent (theoretical) variables. These types of models are purely deterministic.
However, empirical models consist of relationship between measured variables and thus explain the cause and effect relationship between real world economic variables.

Granger (1999) discusses **the empirical model building and the model evaluation** in the context of deforestation in the Amazon region of Brazil. Granger points out the *controversies to construct an empirical model*.

Some econometricians believe in the importance of theory. Some others examine the data and explain relationships between variables.

However, many econometricians use theory to specify the model and then analyse the data in order to find a better representation of the model. Thus, in an empirical modelling the main problem is how to link economic theory and econometric models.

Granger considers a single linear equation with two explanatory variables in order to explain the different statistical properties of competing models.

If the left and the right sides of a model are **balanced,** then the explanatory part of the model can explain the components of the dependent variable.

In other words, *if the dependent variable contains a time trend then the explanatory part must also contain trends to explain the model.*

Therefore, if the dominant property is a trend or a unit root, this property determines the performance of the model.

Granger also explains **the difficulties in model specification**, such as choosing the functional form and lagged orders.
Granger (1999) concludes that consistency and adequacy are important in econometric modelling and a generalized form of relevance is useful for model evaluation.

Overall, the aim in econometric modelling must be to design congruent models with parsimonious specification and mutually low inter-correlations between explanatory variables. So it will be possible to obtain robust models.

Congruent model is the one that has a statistically acceptable representation of the data and it cannot be unambiguously outperformed from any other known model.

Parsimonious specification implies that if two rival models explaining the same phenomenon have the same explanatory power, then the simpler one should be selected.

A model is *robust* in a given time period if small changes, such that extending data set by more variable or more observation, in that period alter the specification marginally, or leave the parameters unchanged.

3.2 Model Specification:

In an econometric modelling, the relationship between the dependent variable and the explanatory variables has been defined either in a form of a **static relationship**, or in a **dynamic relationship.**

A **static relationship** defines the dependent variable as a function of a set of explanatory variables at the same point in time. This form of relation is also called "the long run" relationship. For example

$$y_t = \varphi_1 x_{1t} + \varphi_2 x_{2t} + \varphi_3 x_{3t} + u_t \qquad (1)$$

A **dynamic relation** involves the non-contemporaneous relationship between the variables. This relationship defines "the short run" relationship. Such as

$$y_t = \chi_1 x_t + \chi_2 x_{t-1} + \chi_3 x_{t-2} + u_t \qquad (2)$$

or

$$y_t = \varphi_1 x_{1t} + \varphi_2 x_{2t} + \varphi_3 x_{3t} + \chi_1 x_{1t-1} + \chi_2 x_{2t-1} + \chi_3 x_{3t-1} + u_t \qquad (3)$$

where χ_1, χ_2, and χ_3 are the lag coefficients. The determination of the lag order depends on the response of y_t on a change in x_t. Thus equation (2) and (3) are known as the "**distributed lag models".**

It is assumed that a time series is generated by a stochastic process. However, it is very important to find out which stochastic process will provide the correct data generation process for the series. There are **six types of stochastic process** for the time series modelling, Chatfield (1992). These are:

***1 .A purely random process*:** A time process is called a purely random process if it consists of a sequence of random variables z_t which are mutually independent and identically distributed. That is the process has a constant mean and variance. A purely random process is also called "white noise" process.

***2. Random walk process*:** A z_t process is a purely random process with mean μ and variance σ_z^2, then a process x_t is said to be **random walk** if

$$x_t = x_{t-1} + z_t \qquad (3)$$

The process starts at t=0 and $x_t = \sum_{i=1}^{t} z_i$. The mean and variance of x_t change with t. However, the first difference of a random walk is a purely random process.

3. Moving average process (MA): A time series z_t is a purely random with mean zero and variance σ_z^2, and then a x_t process is called moving average process of order q if

$$x_t = z_t + \beta_1 z_{t-1} + \beta_2 z_{t-2} + \ldots\ldots\ldots + \beta_q z_{t-q} \qquad (4)$$

where β_i are constant. In particular, the MA(1) process is given by

$$x_t = z_t + \theta z_{t-1} \qquad (5)$$

It is important to impose restrictions on β_i in (4) to ensure that the process satisfies the invertibility condition.

By a successive substitution, z_t can be written in terms of x_t. That is

$$z_t = x_t - \theta x_{t-1} + \theta^2 x_{t-2} - \ldots\ldots \qquad (6)$$

If $|\theta|<1$, the model is invertible. The invertibility condition ensures the unique MA process.

MA process is used as an indicator of random events, which will not have immediate but gradual effects on economic variables.

4. Autoregressive process (AR): A time series z_t is purely random process with zero mean and σ_z^2, then the x_t process is defined as an autoregressive process of order p if

$$x_t = \alpha_1 x_{t-1} + \ldots\ldots\ldots + \alpha_p x_{t-p} + z_t \qquad (7)$$

In particular, a first order process AR(1) is given by

$$x_t = \alpha x_{t-1} + z_t \qquad (8)$$

is also called the Markov process.

By successive substitution x_t is defined by the z_t in an infinite order MA process

$$x_t = z_t + \alpha z_{t-1} + \alpha^2 z_{t-2} + \ldots\ldots \qquad (9)$$

where $-1<\alpha<1$. It is required that $|\alpha|<1$ for a stationary process.

5. Autoregressive moving average process (ARMA): An autoregressive moving average process contains p AR terms and q MA terms, and is called an ARMA process of (p,q) order. That is

$x_t = \alpha_1 x_{t-1} + \ldots + \alpha_p x_{t-p} + \beta_1 z_{t-1} + \ldots\ldots + \beta_q z_{t-q}$ (10)

The model can be written in the form

$$\phi(L)\, x_t = \theta(L)\, z_t \qquad (11)$$

where $\phi(L)$ and $\theta(L)$ are the polynomial lag operators of order p and q respectively, such as

$$\phi(L) = 1 - \alpha_1 L - \ldots\ldots - \alpha_p L^p \qquad (12)$$
$$\theta(L) = 1 + \beta_1 L + \ldots + \beta_q L^q \qquad (13)$$

For a stationary process it is required that $\phi(L) = 0$ and $\theta(L) = 0$ to ensure that their roots lie outside the unit circle.

6. Autoregressive integrated moving average process (ARIMA): If the observed time series is not stationary in mean, then it is necessary to remove the non-stationary source of variation. A dth order ARIMA process describes the dth difference of x_t. If d=1, then an ARIMA(0,1,0) process infers the random walk process.

It is generally accepted that a "shock" or "innovation" has a sustained effect in the unit root case, but a diminishing effect with time in the stationary case.

Most of the stationary stochastic processes are purely indeterministic processes. However, a stationary process is written as the sum of two uncorrelated purely deterministic and purely indeterministic processes in the concept of the **Wald decomposition theorem**.

The AR and MA processes are expected to be **purely indeterministic, or stochastic.**

The "deterministic" and "indeterministic" processes can be explained as follows:

Supposing that the series x_t is generated from an AR process, it is possible to decide whether the process is deterministic or indeterministic by the help of estimated variances.

In the estimated linear model, $Var(x_t)$ and $Var(v_t)$ denote the estimated variance of x_t and the residuals (v_t), respectively, thus

$$x_t = \alpha_1 x_{t-1} + \ldots\ldots + \alpha_k x_{t-k} + \nu_t \qquad (14)$$

- If $Var(\nu_t) \leq Var(x_t)$, then $Var((\nu_t)$ is nondecreasing bound as k increases.
- If $\lim_{k\to\infty} Var(\nu_t) = Var(x_t)$, then x_t is purely indeterministic.
- If $\lim_{k\to\infty} Var(\nu_t) = 0$, then x_t is purely deterministic.

In addition, **Wold's decomposition theorem** states that a single stationary time series with indeterministic components having an infinite moving average representation is generally approximated by a finite autoregressive moving average. If the data generating process for y_t is

- First order autoregressive process AR(1):

$y_t = \rho\, y_{t-1} + u_t$,

where $u_t \sim iid\,(0, \sigma^2)$ and $|\rho| < 1$, AR(1) is always stationary.

- First order moving average process MA(1):

$y_t = u_t + \zeta_1 u_t$

where $u_t \sim iid\,(0,\sigma^2)$, MA(1) is always stationary.

- First order autoregressive moving average ARMA(1,1):

$y_t = \rho\, y_{t-1} + u_t + \zeta_1 u_t$

The **London School of Economics (LSE) methodology** is a mid-point between the classical econometric methodology and the theoretical pure time series analysis. This methodology is based on the **"general to specific" modelling approach** discussed by Hendry and Mizon (1978), Hendry (1980), and Mizon and Richard (1986).

The general to specific approach has had an important influence on applied econometrics.

This approach is carried out to select a model that is consistent with economic theory and captures the data generating process.

A general to specific model selection starts from a very general model that is determined according to relevant economic theory, available data and acceptable diagnostic tests results.

This model selection strategy makes initial **over-parametrisation** deliberately and aims to reduce the risk of inconsistency.

3.3 The Data Generating Process:

Time series theory starts with the **data generating process (DGP)** for the series.

The DGP determines the properties of the data set which influences the results obtained from any given relationship by a specified method. In other words, this process should be able to generate all the statistical properties of the series conditional on past data. Some series appear to be "stationary" which in fact implies that the linear properties are available and they are time independent.

"General to Specific" approach is an econometric methodology, and is based on the data generating process concept. The DGP represents a general statement of the joint probability distribution of all variables. The general data generating process stated as the joint probability of the sample x_t is

$$\prod_{t=1}^{T} D(x_t \mid X_{t-1}, \lambda) \qquad (15)$$

where x_t is a vector of observations on all variables in period t, and

$X_{t-1} = (x_{t-1}, \ldots, x_1)'$, λ is a vector of unknown parameters.

An empirical model must be the "reduction" of whatever process generated the observed data. This depends upon the constraints imposed by the theory and the properties of the underlying data generating process. Gilbert (1986) presents an excellent discussion on the general to specific modelling approach and the data generating process.

The process of econometric modelling simplifies the DGP by assuming general steps:

(i) *Marginalising* the DGP. Since the DGP contains many variables, there should be a selection of "variables of interest" and elimination of some others.
(ii) *Conditioning* assumption. Given the variables of interest, endogenous variables (y_t) should be chosen and then these variables should be conditioned or determined by the remaining variables (z_t). The z_t variable should be at least weak exogenous for a valid conditioning.
(iii) *Selection of functional form*. A specific functional form should be selected. This form should be a suitable simple specification.
(iv) *Estimation*. Unknown parameters should be replaced by estimated values.

In the first three steps economic theory could be utilized, then econometric theory should be used in the last step.

Considering the above steps equation (15) can be reparameterized as the following:

$D(x_t | X_{t-1}, \lambda) = A(y_t | Y_{t-1}, z_t ; \lambda_1)\ B(z_t | Y_{t-1}, Z_{t-1}; \lambda_2)$ (16)

where $A(y_t | Y_{t-1}, z_t; \lambda_1)$ specifies the **endogenous variables** of interest **as a function of lagged endogenous variable and exogenous variable,** and

$B(z_t | Y_{t-1}, Z_{t-1}; \lambda_2)$ gives the **determination of the exogenous variables as a function of the lagged endogenous and exogenous (predetermined) variables.**

Engle, Hendry and Richard (1983) discuss that in the econometric modelling certain variables can be treated as fixed for the statistical purposes. This is called "weak exogeneity". In addition, it is possible to break the process up into independent parts for understanding and controlling a process. This is called "super exogeneity".

Equation (16) represents the restricted form of a very general statement such that reduction and elimination of some variables, and transforming and treating some variables given.

Accordingly, y_t is the endogenous variables that are determined or conditioned by the remaining variable z_t. The y_t and z_t are the variables of interest. The **z_t should be, at least, weakly exogenous or more simply z_t must be independent of y_t.** It is also required that the parameters of interest of the model, λ, be estimated as a function of λ_1 only, **and λ_1 and λ_2 are variation free.**

If the second term of equation (16) is expressed as $B(z_t | Z_{t-1}; \lambda_2)$, then **the exogenous variables are determined without any effect of lagged endogenous variables**. In this case z_t is "strongly exogenous". That is, the strong exogeneity is given by the assumptions of weak exogeneity and Granger non-causality.

If z_t is determined without any effect of lagged endogenous variables, then it is known that "Y_{t-1} does not cause z_t". This is defined as the **Granger non-causality**.

Another form of exogeneity is the super exogeneity, and it requires strong exogeneity and the assumption of independence of λ_1 and λ_2.

In general, weak exogeneity is necessary for model estimation and testing, strong exogeneity is essential for forecasting, and super exogeneity is required for policy analysis.

3.4 Estimation of a Single Equation:

The primary goals of an econometric modelling are: first, estimating the coefficients of interest and forecasting; second, predicting the movements in the dependent variable. If the estimated model fits the data well and satisfies the assumptions about the error term, then the estimated equation will explain the variations of the dependent variable in the sample.

In general each single equation is estimated by Ordinary Least Squares (OLS) method. The principle of the OLS technique is estimating the coefficients of an equation *to minimize the residual sum of squares*.

The assumed model is $y_t = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + u_t$, then the least squares residuals (e_t) is defined as the difference between the actual (y_t) and estimated values ($a_0 + a_1 x_{1t} + a_2 x_{2t}$). That is

$$y_t - a_0 - a_1 x_{1t} - a_2 x_{2t} = e_t \qquad (17)$$

where a_i are the estimators of the parameters α_i.

The conventional econometric modelling techniques assume that *the regression equation is linear in coefficients, the error terms are identically, independently and normally distributed with a zero mean and constant-finite variance, explanatory variables are independent of each other and they are uncorrelated with the error term*. As long as the assumptions are valid, the OLS estimators are accepted to be best linear unbiased estimator (BLUE) as proved by the *Gauss-Markov theorem.*

Assuming that the economic time series data is generated by an **autoregressive distributed lag model, ARDL(p, q),** where p and q are the lag lengths of the dependent and independent variables respectively, then consistent and asymptotically efficient estimates can obtained by OLS method. The **ARDL(p,q)** model is specified as

$$y_t = \alpha_0 + \alpha_1 x_{t-1} + \ldots + \alpha_p x_{t-p} + \beta_1 y_{t-1} + \ldots + \beta_q y_{t-q} + \varepsilon_t \qquad (18)$$

The reason in including the lagged values of the explanatory variables and dependent variables is to estimate the dynamic relationship between economic variables. The cost of including lagged variables in the estimation process is the loss of degrees of freedom and the degree of multi-collinearity.

It is usually suggested that **the lag structure of a model depend upon the time units of data.** If the model is specified by using annual data one lag must be applied for each variable. Respectively, four lags must be applied for quarterly data, and twelve lags for

monthly data. Thus the choice of p and q depends solely on the data constraint and also the data frequency.

In addition, the number of exogenous and endogenous variables in a single equation is limited according to the number of observations.

As suggested by **Sargan** (1980), the maximum number of variables in an equation must be equal to two and a half times the number of observations in order to eliminate small sample bias. The constant term should always be included in a regression model among exogenous and lagged endogenous variables because it represents the origin of the estimated empirical relationship.

In the general to specific modelling technique, the specific form of an equation is obtained by considering the significance of each estimated coefficient according to the t-values. The coefficients whose t-values are less than the tabulated values are eliminated successively, starting with the lowest one.

The statistic for the goodness of fit is the coefficient of determination, R^2. *It measures the proportion of the variation in the dependent variable explained by the explanatory variables.* R^2 is a useful statistic for evaluating the fit of the estimated model and comparing models with different data sets.

Even though R^2 provides information about the explanatory power of the equation, *it always increases as new variables are added to a regression model since it does not take into account the degrees of freedom.*

However, adjusted R^2 is a more desirable measure since its value may increase or decrease when new variables are added to the equation.

The F-statistic is used in the regression equation to test the significance of the R^2 statistic, hence the overall significance of the estimated equation.

All these statistics are used to judge the significance of the estimated coefficients. However, **rigorous evaluation** of the model is achieved by diagnostic checking of residuals.

The diagnostic tests are used for checking the serial correlation, heteroskedasticity, normality and functional form.

Serial correlation is defined as *the correlation of the residuals*. This is a violation of the assumption that the residuals are uncorrelated with each other.

Serial correlation **can arise** due to *omission of relevant variables, using incorrect functional form, inertia in economic time series or data manipulation*.
In time series data, the error term of a specific time period is likely to be determined systematically by the previous time periods' error terms.

Serial correlation affects the precision, but not the accuracy of the estimated coefficients in a model. The estimates are unbiased, but their true variances are underestimated when serial correlation is present. As a result, the sum of squared residuals may underestimate the true unexplained variations, causing the t-values and F-values to be higher than they should otherwise be. Hence some coefficients are likely to appear statistically significant when in fact they are not.

Heteroskedasticity is present if *the variance of each error term conditional on the chosen values of explanatory variables is not constant over the sample period.* That means unequal error variance.

Heteroskedasticity **can arise** mainly because of presence of outliers in the sample data. Heteroskedasticity does not destroy the unbiasedness and consistency properties of the OLS estimators, but affects efficiency. Hence, in the presence of heteroskedasticity the OLS estimators are no longer the best, and t and F statistics would be misleading.

Normality assumption of the error terms ensures that the OLS estimators for the partial slope coefficients are also normally distributed. The OLS estimators would be unbiased, efficient, and consistent if the normality is accepted. Skewness and kurtosis coefficients are used to check the normal distribution. On the other hand, functional form is closely related to the linearity assumption.

In addition to these diagnostic tests, it is also necessary to check the presence of a structural break date, or parameter constancy, within the sample period. **Chow structural break test**, cumulative sum (CUSUM) and cumulative sum of squares (CUSUMSQ) techniques detect structural breaks. The CUSUM plot is applied with recursive residuals that behave very different from the OLS residual in a misspecified model, Harvey (1981). If the model is misspecified, there would likely be a tendency for disproportionate number of recursive residual to have the same sign. The cumulative effect of this will tend to move away from the horizontal axis. On the other hand, the CUSUMSQ plot provides a complement to the CUSUM plots and is used with OLS and recursive residuals.

These tests provide **further information about the robustness of the regression model to changes in the data.**

3.5 ARCH and GARCH Models:

In traditional econometric models, the variance of the disturbance term is assumed to be constant (homoskedasticity).

However, many economic time series fluctuate during some time periods and grow smoothly during some other periods. ***Many economic time series exhibit large volatility over time.*** These series are called conditional heteroskedastic. This is a kind of heteroskedasticity in which the variance of the error term depends upon the volatility of the errors in the past.

The ARCH model was initially proposed by **Engle in 1982** and then modified by others. GARCH model was developed by **Bollerslev in 1986**.

The autoregressive conditional heteroskedastic, ARCH(p), and the generalized ARCH, GARCH(q,p), models can be explained in the context of an ARMA model, Enders (1995).

Consider an ARMA model $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$ without a constant variance. The conditional variance of y_{t+1} is

$$\text{Var}(y_t \mid y_{t+1}) = E_t[(y_{t+1} - a_0 - a_1 y_t)^2] = E_t\varepsilon^2_{t+1} \qquad (19)$$

Since $E_t\varepsilon^2_{t+1} \neq \sigma^2$, then the conditional variance can be modelled as an AR(p) process using the square of estimated residuals:

$$e_t^2 = a_0 + a_1 e_{t-1}^2 + a_2 e_{t-2}^2 + \ldots\ldots + a_p e_{t-p}^2 + u_t \qquad (20)$$

where u_t is a white noise process.

If the values of $\alpha_1, \alpha_2, \ldots\ldots, \alpha_n$ all are equal to zero, then the estimated variance will be α_0. Otherwise, the conditional variance at t+1 is obtained by

$$E_t e_{t+1}^2 = a_0 + a_1 e_t^2 + a_2 e_{t-1}^2 + \ldots\ldots + a_p e_{t+1-p}^2 \qquad (21)$$

Equation (20) is called an ARCH model. There are many possible applications for ARCH models. In equation (20) the residuals can be obtained from an ARMA model, or an autoregression model, or a traditional regression model.

If the conditional variance is generated by a first order autoregressive process, ARCH(1), then the coefficients α_0 and α_1 should be **positive** to ensure that the conditional variance is positive. Moreover, to ensure the stability of the process, it is necessary to restrict α_1, such that $0< \alpha_1<1$.

In an ARCH model, the errors are serially uncorrelated, but they are related by means of their variances. The conditional variance itself is an autoregressive process and it results in conditionally heteroskedastic disturbances. ARCH models can capture the volatility within the univariate framework.

Franses (1998) discusses conditional heteroskedasticity and some properties of time series data that are generated by ARCH and GARCH models.

An ARCH(1) model can describe a time series with outliers. The presence of the sequence of outliers can be observed by (i) autocorrelation in y_t^2 series or in the squared residuals, (ii) excess kurtosis, and (iii) negative or positive skewness.

The GARCH(q,p) model includes **both the information about volatility observed in the previous period, i.e. short run volatility, (ARCH term) and the forecasted variance from last period, i.e. long run volatility, (**GARCH term) in order to predict the current period's variance.

Thus GARCH models describe both the autoregressive and moving average components of time series data with the heteroskedastic variance.

The GARCH(q,p) model is

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2 \qquad (22)$$

If q=0 and p=1, then the process is the first order ARCH model.
If q=1 and p=1, then the GARCH(1,1) model must be estimated in order to make the correct inference.
However**, if α+β=1,** then volatility shocks will be persistent. In equation (22) all the coefficients must be positive.

The **autocorrelation function (ACF)** and **partial autocorrelation function (PACF) of the residuals** can be helpful in order to identify the order of the ARCH and the GARCH models. The ACF of the squared residual is considered to identify the order of the GARCH model.

3.6 Estimation of System Equations:

A simultaneous model shows a relationship between two or more endogenous variables in terms of other endogenous variables, predetermined variables and disturbances.

A model with more than two jointly dependent variables requires a distinction between the dependent and independent variables.

Jointly dependent variables are defined as a whole by the model. These variables are called "endogenous" variables. However, some of the variables are determined outside the model. These are called "exogenous" variables. "Predetermined" variables are defined as a group of current and lagged exogenous variables and lagged endogenous variables.

It is important to obtain consistent and efficient estimates of the coefficients of interest. The estimation results are mainly determined by the estimation method. System estimations should be preferred to single estimation techniques because of the lower variance of estimates. However, the difficulty with the system estimation techniques is that individual coefficient estimates are sensitive to the specification of the model. A specification error in one equation may affect other estimators. Hence, using simultaneous estimation method is likely to produce both the gain of efficiency and the cost of specification error.

The appearance of endogenous variables as explanatory variables in the system equations results in correlation between the explanatory variables and the disturbances. This violates the independency assumption of the linear model and hence application of the OLS technique yields biased and inconsistent estimators. In order to provide consistent estimates and the contemporaneous feedbacks of the equations, the instrumental variables methods, such as the two stage least squares (2SLS), or the three stage least squares (3SLS), or the full information maximum likelihood (FIML) methods, are used to estimate the system equations.

4. STATIONARITY VS NONSTATIONARITY

4.1. Definition of Stationarity in Time Series:

A primary consequence of empirical models being reduction of a data generating process is the "error term".

*The **error term** on an econometric model is a generated process and it is expected to be loosely associated with exogenous shocks to an economy.* The error must represent **"everything not elsewhere specified"** in the specified equation. Thus it varies with the equation formulation and it must be designed to satisfy certain criteria, such as stationarity, Hendry and Mizon (1985).

The **stationarity of the error term** means that the time series y_t does not deviate systematically from their theoretical determinants.

The stationarity of the error process is a substantial test for the validity of the long run equilibrium or economic theory.

In the economic theory certain pair of economic variables should not diverge from each other remarkably in the long run or at the steady state, however they may drift apart in the short run.

The classical linear regression models are based on the assumptions that the mean, variance and autocovariance are finite, constant and time invariant. In other words, in econometric modelling, the time series are generated by a stationary stochastic process.
A "stochastic process" is a sequence of "random variables", any realisation of a stochastic process is a "time series", any element of a time series is an "observation".

Consider a stochastic process for y_t that is defined for t=….,-1,0,1,… . The y_t is called **"covariance (or weakly) stationary",** if

(i) the mean (μ) and the variance (σ^2) of y_t are constant, finite and time independent,
(ii) covariance (σ^2_τ) between two values of y_t depends only upon the distance (time interval (τ)) between those two values, but not upon the point in time t, then.

That is:

$$E(y_t)=\mu \qquad (1.1)$$
$$V(y_t)=E[(y_t-\mu)]=\sigma^2 \qquad (1.2)$$
$$Cov(y_t,y_{t-\tau})=E[(y_t-\mu)\,(y_{t-\tau}-\mu)] = \sigma^2_\tau \qquad (1.3)$$

Equation (1.1) to (1.3) define the mean, variance and autocovariance, which are unknown "population" values. From any given realisation of the stochastic process, generating the data, we have the sample values of these population values. An adequate illustration of this discussion is given by the following statistics:

$$\bar{Y} = \frac{\sum_{t=1}^{T} y_t}{T} \qquad (2.1)$$

$$s^2 = \frac{\sum_{t=1}^{T} (y_t - \bar{Y})^2}{T} \qquad (2.2)$$

$$s_\tau^2 = Cov(y_t, y_{t-\tau}) = E\left[\left(y_t - \bar{Y}\right)\left(y_{t-\tau} - \bar{Y}\right)\right] \qquad (2.3)$$

These statistics present consistent estimates of the population mean, variance, and autocovariance. If the sample autocovariances are standardised dividing by the sample variance, we obtain the **sample autocorrelation coefficient, r.** That is:

$$r^\tau = s_\tau^2 / s^2 \qquad (3)$$

The graph of the sample autocorrelation against the distance between two values of time series y_t is known as the **"correlogram".**

Although correlogram provides us visual information about the stationary of the series, it is not a formal test.

Consider a pure time series model which is generated by a univariate autoregressive process of order one, i.e. AR(1).

$y_t = \rho\, y_{t-1} + u_t \qquad t=\ldots..,-1, 0, 1, \ldots.. \qquad (4)$

where u_t is assumed to be a sequence of independently and identically distributed (i.i.d.) random variable with expected value zero and variance σ^2, i.e. $u_t \sim iid\ (0, \sigma^2)$.

The process starts at some point in the distance past and ρ is a real number. **The coefficient of y_t, ρ, measures the degree of persistent deviations of y_t from its mean.**

The data generating process in (4) is:

- Stationary, when **ρ is less than one** in absolute value, i.e. $|\rho| < 1$.

Thus the series returns to its mean and fluctuate around it with a constant range.

- Non-stationary, when **ρ is equal to one**, i.e. ρ= 1.

This implies that deviations are permanent and y_t has a **"unit root".** The series has a different mean at different points in time, variance increases with the sample size.

If $|\rho| < 1$, the AR(1) process can be presented as a moving average process of infinite order, i.e. MA(∞). Given the assumptions about the error term in equation (4), we obtain the following results:

$E(y_t)=0$ (5.1)

$V(y_t)= \sigma^2 (1-\rho)^{-1}$ (5.2)

$Cov(y_t , y_{t-\tau}) = \sigma^2 \rho^{\tau} (1-\rho)^{-1}$, $\Gamma= 1,2,..$ (5.3)

$Corr(y_t, y_{t-\tau}) = \rho^{\tau}$ (5.4)

The statistics from (5.1) to (5.4) show that the mean, variance and covariance do not depend on time t, and the correlation coefficient ρ^{τ} is expected to fade away rapidly as the lag length τ increases. Accordingly, **the AR(1) process is a stationary process.**

Instead, the process in (4) has a unit root process if and only if ρ is one. In this case, the time series is non-stationary implying that the variance of y_t increases with time, i.e. $V(y_t)=t\sigma^2$, without bound and any correlation coefficient ρ^{τ} remains constant at 1 as τ increases, i.e. $\rho^{\tau} =1$.

Thus the Gauss and Markov theorem stating the minimum variance property of all linear unbiased estimators does not hold, and the conventional methods yield inconsistent parameter estimators.

A visual inspection of a correlogram can provide information about the data generating process for the time series. The estimated autocorrelation coefficients fade away rapidly as the lag length increases for a stationary series, but they decline slowly for a non-stationary series.

4.2. Stochastic and Deterministic Time Trends:

In the econometric literature it is widely argued that most of the economic variables have changed in mean and variance. On this basis, it is accepted that the conditional distribution of the regression is not stationary. An important evidence of non-stationarity is the high serial correlation between successive observations.

Some econometricians have suggested differencing and pre-whitening[1] the time series prior to estimating models in order to overcome the non-stationarity problem. It would be better to say that the series must be transformed in order to attain stationarity. A common way of achieving stationarity is to take logarithms and then difference.

Granger (1986) discusses that some series appear to be stationary provided that the time series have the linear properties and are time invariant. Nonetheless, some series need to be differenced to achieve the linear properties of the series.

Granger defines **"integration of series".** If a time series y_t needs differencing d times to become stationary, it is called "integrated of order d" and denoted as $y_t \sim I(d)$.

On the basis of this definition, if a series y_t is stationary such series is called "integrated of order zero", denoting $y_t \sim I(0)$.

If a series y_t is non-stationary and need to be differenced once to achieve the stationarity, such series is called "integrated of order one", denoted $y_t \sim I(1)$.

A white noise error term[2] is the simplest example of an I(0) series.

A stationary AR(1) series where y_t is generated by $y_t = \rho\, y_{t-1} + u_t$ with $|\rho| < 1$ and u_t is white noise with zero mean is an another example of an I(0) series.

On the other hand, the simplest example of an I(1) series is a "random walk", where y_t is generated by $y_t = \rho\, y_{t-1} + u_t$ with $\rho = 1$. In that case the first difference series is white noise, i.e. $\Delta y_t = u_t$.

---

[1] Pre-whitening procedure is known as the removing of trend and seasonality from the time series.

[2] A white noise error term is defined as a random variable which is independently and identically distributed with a zero mean, constant variance, and zero correlation for all non-zero lags. A white noise error term would not exhibit systematic errors. For this reason in econometric investigations diagnostic tests for residual should be applied to confirm the stationarity of residuals.

Granger points out some essential differences between I(0) and I(1) series:

- An I(0) series has a constant mean and the series tends to return to the mean provided that it fluctuates around the mean value. Autocorrelation declines rapidly as lag increases and the process gives low weights to events in the medium to distant past and thus has a finite memory.
- An I(1) series without drift will be relatively smooth, wander widely and rarely return to an earlier value. Autocorrelations are near one in magnitude even for large samples and the process has an indefinitely long memory.

Nelson and Plosser (1982) explain the decomposition of economic variables as the **secular component and the cyclical component.**

They define the secular component in the context of growth theory with real factors, but the cyclical component is transitory with monetary factors.

Since cyclical fluctuations are assumed to disperse over time, any long run or permanent movement is attributed to the secular component.

As the secular component moves slowly and smoothly relative to the cyclical component, secular component needs to be modelled by a deterministic trend.

Accordingly, the form of non-stationarity could be classified on the basis of **a stochastic trend and a deterministic trend in macroeconomic time series.**

In the least square regression analysis the inclusion of a time trend among the regressors is practically the same as the use of **detrended time** series, i.e. regressing y_t on a time trend and obtaining residuals. The residuals are interpreted as the *cyclical component*.

It is important to explain a random walk process that exhibits secular movement, but does not follow a deterministic path.

Overall, if the secular movement in macroeconomic time series is of a stochastic rather than deterministic, then an inclusion of a time trend as an explanatory variable in the model will result in misspecification.

However, if the secular component moves slowly and smoothly relative to the cyclical component, this component needs to be modelled by a deterministic trend.

Considering the detrending procedure, the model is defined as:

$y_t = \alpha+\beta t + u_t$, and (6)

$\phi(L)\, u_t = \theta(L)\, \varepsilon_t$ (7)

where y_t is in natural logarithm form, t is a time trend and u_t is white noise disturbance term; α and β are fixed parameters; $\phi(L)$ and $\theta(L)$ are polynomials in L that satisfy the condition for stationarity and invertibility; $(\alpha + \beta t)$ is trend component and u_t is nontrend component; **β measures the per period growth rate of y_t.**

It is assumed that non-trend component is a stationary stochastic process with zero mean, whereas trend component is a deterministic function of time. Thus, y_t is said to be a **"trend stationary"** variable, it may have a trend component but its deviation from a deterministic trend is stationary.

An alternative model is:

$\Delta y_t = \beta + u_t$, and (8)

$\delta(L)\, u_t = \lambda(L)\, \upsilon_t$ (9)

where u_t is assumed to be stationary and $u_t = \upsilon_t$ for a random walk process; $\delta(L)$ and $\lambda(L)$ are polynomials; Δy_t is the first difference, **β is a measure of the average per period change in y_t.**

In (8) y_t is unit root non-stationary, it has a stochastic trend, and is generated by a **"difference stationary" process**.

A non-stationary process y_t becomes stationary and invertible autoregressive moving average (ARMA) process by differencing first or higher order. Thus y_t is called **"difference stationary"** variable.

Nelson and Plosser explain the fundamental difference between the trend stationary and difference stationary processes by expressing y_t as the value at some reference point in the past, time zero plus all subsequent changes,

$$y_t = y_0 + \beta t + \sum_{j=1}^{t} u_j \qquad (10)$$

Equations (6) and (10) show that the two types of processes can be written in terms of a linear function of deterministic time trend, and also the deviations from trend. It can be seen that

- The intercept term in (6) is a fixed parameter, while in (10) y_0 is a linear function of historical events.
- The deviations from trend in (6) are stationary, whereas in (10) they are accumulation of stationary changes. The accumulation in (10) is not stationary, but its variance increases without bound as t gets larger.
- The long run forecast error in (6) has a finite variance, but in (10) the variance of the forecast error will increase without bound.
- The two processes can be written in terms of the roots of the AR and MA polynomials.

➢ If the linear trend stationary process is differenced once, the result is

$$\phi(L)\,[(1-L)y_t] = \beta\,\phi(L=1) + (1-L)\,\theta(L)\,u_t \qquad (11)$$

where $\phi(L=1)$ is a constant obtained by evaluating the polynomial $\phi(L)$ at L=1, $[(1-L)y_t]$ the first difference operator.

Equation (11) displays that a unit root will be present in the MA component of the ARMA process. In this case the presence of the unit root process reveals that the process is not invertible and cannot have an AR representation. However, the difference stationary series are both invertible and stationary.

➢ If the difference stationary process is expressed in terms of levels, the outcome is

$$\delta(L)(1-L)\,y_t = \beta\,\lambda(L=1) + \lambda(L)u_t \qquad (12)$$

Equation (12) contains unit root in the AR polynomial.

Accordingly, the important result is that **difference stationary variable cannot be made stationary by the removal of a deterministic trend.** Instead it should be differenced. Thus the difference stationary process is purely stochastic, whereas the trend stationary process is deterministic.

Differencing approach, however, neglects the important information about the long run relationship between the levels of the series.

Hylleberg and Mizon (1989) argue that it is not necessary to model in the differences of variables because it is necessary that all variables included in the regression be stationary, in fact, it is necessary that the conditional distribution be stationary.

Molinas (1986) considers a moving average process rather than the autoregressive one. Thus he assumes that a random walk variable is not determined by the sum of all the past innovations, but by the error dynamics. He defines the regression as

$y_t = \alpha + \rho\, y_{t-1} + u_t + \zeta_1 u_{t-1}$ (13)

where $\rho=1$, $u_t \sim$ iid $(0, \sigma^2)$. If $\zeta=0$, y_t is a random walk, or if $\zeta=-1$, y_t is white noise.

Molinas argues that if $|\zeta| < 1$, the mean (i.e. $\alpha \neq 0$) and variance of y_t increases with t, but if $\zeta > 1/2$, the correlogram of y_t in fact dies out rapidly, unlike the definition of an I(1) series in Granger's paper.

Campbell and Mankiw (1987) define a measure of the ultimate or long run response of y_t to a unit shock. Consider a difference stationary process:

$\Delta y_t = \beta + u_t$ (14)

$\Delta y_t = \beta + A(L)u_t$ (15)

where A(L) lag polynomials in L,

Thus

$y_t = y_{t-1} + \beta + u_t + A_1 u_{t-1} + A_2 u_{t-2} + \ldots$ (16.a)

$y_t = y_{t-1} + \beta + A(1)\, u_t$ (16.b)

Equation (16.a) implies that per unit effect of u_t on y_t is one. The value **of A(1) varies with the response pattern,** such as

A(1) > 1, if the ultimate impact of a shock on y_t is greater than the first period impact.

A(1) = 1, if the data generating process is a random walk process.

A(1) = 0, if the data is generated from a trend stationary process.

4.3. Tests for Stationarity:

The concept of the unit root has provided the evidence that a large number of economic variables do not deviate temporarily from stable long run equilibrium, yet the deviations are permanent.

As a consequence, an application of classical estimation methods to regression gives misleading inferences.

In econometric modeling, presence of common trends in the data leads to spurious correlations between variables and hence implies a relationship between the variables, even though each series has a different reason to grow at different rates.

The non-stationary nature of the series results in a kind of causal relationship and the chance of **"spurious regression"**.

In the regression with difference stationary variables without drift, the random walk components will dominate the stationary components, and lead the t-statistic on the parameters of interest to be considerably larger than the critical values.

Yule (1926) argues that in a regression model with unrelated non-stationary variables the R^2 tends to unity.

In addition, Granger and Newbold (1974) suggest that the rule **R^2 > Durbin-Watson (DW)** statistic could be used in order to understand and recognise the spurious regression.

Phillips (1986) shows a different implication of the spurious regression problem and proposes that a low DW statistic could often indicate the non-stationarity of the variables.

A high R^2 may indicate correlated trends and not true equilibrium relationships, but low Durbin-Watson statistic may provide information about the non-stationary residuals.

DW statistic is used to test if the error terms are correlated or not. The DW statistic is based on the residual from the estimated regression model assuming there is not any lagged value of the dependent variable.

However, if the generating process of error terms is determined not only by its sequence but also by the sequence of all the statistically significant independent variables, then the lagged value of the dependent variable appears among the explanatory variables. In this case the Durbin's h statistic is the appropriate statistic for error correlations. Therefore, the DW statistic cannot always give a correct interpretation of the estimation results.

The investigation of stationarity in a time series is closely related to the tests for unit roots.

Tests For the Null of Difference Stationarity:

a- The Dickey and Fuller Tests
b- The Augmented Dickey and Fuller, (or Said and Dickey), Tests
c- The Phillips and Perron Tests
d-The Park J Tests

a- The Dickey and Fuller (DF) Tests:
Dickey and Fuller (1979, 1981) consider the AR(1) process in

$$y_t = \rho\, y_{t-1} + u_t \qquad (17)$$

The testing procedure is based on the assumption that ***u_t is an identically and independently distributed random variable with zero mean and constant variance σ^2.***

In testing for unit roots, we are essentially testing the null hypothesis

H_0: $\rho = 1$ against H_1: $|\rho| < 1$.

Under the null hypothesis y_t is non-stationary, it is sometimes called a random walk without drift. Under the alternative hypothesis y_t is a stationary AR(1) process.

The re-parameterisation of (17) is obtained by subtracting y_{t-1} from both sides of the equation, that is:

$$y_t - y_{t-1} = \phi\, y_{t-1} + u_t \qquad (18)$$

where $\phi = (1-\rho)$. Then it would be possible to test H_0: $\phi = 0$ against $H_1 = \phi < 0$. The conventional statistic for ϕ in (18) is identical to the t-ratio for $\rho = 1$ in (17).

Dickey and Fuller (1979) investigate three models for the data generating process for y_t.

$y_t = \rho y_{t-1} + u_t$, $t=1,2,\ldots$ and $y_0 = 0$ (19.1)
$y_t = \alpha + \rho y_{t-1} + u_t$, $t=1,2,\ldots$ and $y_0 = 0$ (19.2)
$y_t = \alpha + \beta t + \rho y_{t-1} + u_t$, $t=1,2,\ldots$ and $y_0 = 0$ (19.3)

Note that tests for the null hypothesis of a unit root will not be based on standard t distribution, the critical values should be obtained from tables in Fuller (1976). Dickey and Fuller (1979) derived a non-standard limiting distribution for the least square t statistic for the H_0: $\rho = 1$.

When (19.1) is the estimated equation, the test statistic for $\rho=1$ is Γ. When (19.2) is the estimated equation, the test statistic for $\rho=1$ is Γ_μ. When (19.3), which is called "Dickey and Fuller regression", is the estimated one the test statistic for $\rho=1$ is Γ_Γ.

The limiting distribution of Γ_Γ is derived under the assumption that the coefficient for time trend is zero, and the distribution is not affected by the value of α.

The data generating process for y_t would be different for each specified equation, such that

- If $|\rho| < 1$, then
  in (19.1) is a zero mean stationary AR(1) process
  in (19.2) is a stationary AR(1) process with a drift, if $\alpha \neq 0$
  in (19.3) is a stationary AR(1) process with a drift and deterministic trend, if $\alpha \neq 0$, $\beta \neq 0$.

- If $\rho = 1$, then
  in (19.1) y_t is I(1) and is random walk without drift.
  in (19.2) y_t is I(1) and is random walk with drift, if $\alpha \neq 0$
  in (19.3) y_t is I(1) and is random walk with drift and a time trend, if $\beta \neq 0$, $\alpha \neq 0$,

- If $|\rho| > 1$, then
the time series is not stationary and the variance of the series grows exponentially as time increases.

Notice that the appropriate testing procedure will depend upon what is chosen for maintained model, null hypothesis and the form of alternative hypothesis. Whether or

not a constant and/or a linear time trend should be included in the regression depends on what type of alternative is appropriate.

- If the H_1 is that y_t is stationary with mean zero, then no deterministic terms should be included. This hypothesis is not appropriate for most of the macroeconomic time series.

- If the H_1 is that y_t is stationary with unknown mean, then a constant should be included. This hypothesis is appropriate for ***the time series that exhibit persistent growth over time***.

- If the H_1 is that y_t is trend stationary, then a constant and a linear trend should be included. This hypothesis is appropriate for ***the time series which shows a consistent tendency to grow over time***.

When the test statistics are negative and greater than the appropriate critical values (in absolute values), then the null of a unit root will be rejected in favour of one of the alternatives.

Holden and Perman (1994) indicate the possible outcomes of the maintained model (19.3) under the null hypothesis:

Case.1 : $\beta = 0, \alpha = 0$
Case.2 : $\beta \neq 0, \alpha = 0$
Case.3 : $\beta = 0, \alpha \neq 0$
Case.4 : $\beta \neq 0, \alpha \neq 0$

Case.1 implies that model (19.1) is correct, but model (19.2) includes an unnecessary intercept term, model (19.3) includes both unnecessary intercept and time trend. Case.2 shows that model (19.1) excludes a necessary time trend, model (19.2) includes an unnecessary intercept term and excludes a required time trend, and model (19.3) includes an irrelevant intercept term.

Note that since the alternative hypothesis is $|\rho|<1$ in each case, the null hypothesis of non-stationarity could be rejected in favour of the alternative of stationarity for the negative and smaller values of test statistic.

In the estimation process it is important to decide whether or not β is zero. Single coefficient tests for the presence of unit root from estimating different alternative combinations will not be sufficient to know the correct values of α and β. Therefore, the joint coefficient tests for the presence of a drift and/or a deterministic trend besides a unit root will improve the testing results.

Dickey and Fuller (1981) construct the likelihood ratio statistic (LR) for the null hypothesis that the maintained model is a random walk with drift. They derived test statistics for the different cases. These are

	Test statistic	Estimated equation
Case.1 H_0: $(\alpha, \rho) = (0,1)$	Φ_1	(19.2)
H_1: $(\alpha, \rho) \neq (0,1)$		
Case.2 H_0: $(\alpha,\beta,\rho) = (0,0,1)$	Φ_2	(19.3)
H_1: $(\alpha,\beta,\rho) \neq (0,0,1)$		
Case.3 H_0: $(\alpha,\beta,\rho) = (\alpha,0,1)$	Φ_3	(19.3)
H_1: $(\alpha,\beta,\rho) \neq (\alpha,0,1)$		

Test statistics **Φ_1, Φ_2, Φ_3 are the common regression F tests** and they are calculated by the restricted and unrestricted residual sums of squares. The distribution of the test statistic Φ_3 is not affected by the presence of the intercept term. The F test rejects the null hypothesis for the large values of test statistics.

Perman (1991) demonstrates the testing procedure for unit roots and stationarity considering several models. Perman considers the models (19.1), (19.2), and (19.3), and a sequence of tests for stationarity.

- If the maintained model is $y_t = \rho\, y_{t-1} + e_t$, then

the null hypothesis $H_0 : \rho = 1$, or $\Delta y_t = e_t \Rightarrow$ a random walk with no drift

the alternative hypothesis H_1: $-1<\rho<1$, or $\Delta y_t = \rho\, y_{t-1} + e_t$, $\rho \neq 0 \Rightarrow$ a stationary AR(1) model.

Using the test statistic Γ, the null hypothesis will be accepted if test statistic is greater than the tabulated critical values in Fuller (1976). Note that the use of test statistic is only valid for the time series with zero mean value.

- If the maintained model is $y_t = \vartheta + \rho\, y_{t-1} + e_t$, where $\vartheta = \mu\,(1 - \rho)$, then

the null hypothesis $H_0 : \rho = 1$, or $\Delta y_t = e_t \Rightarrow$ a random walk with no drift

the alternative hypothesis $H_1 : -1 < \rho < 1$, or $y_t - \rho\, y_{t-1} = \mu\,(1 - \rho) + e_t$.

Note that under the null hypothesis $\mu\,(1-\rho) = 0$ and thus the first difference of y_t is a random walk with no drift. If stationarity is tested under the null hypothesis the

variable should be in first difference, whereas the variable should be in the levels if stationarity is tested under the alternative hypothesis.

The alternative hypothesis could be written as $(y_t - \mu) = \rho\ (y_{t-1} - \mu) + e_t$, ρ measures the degree of persistence of deviations of y_t from μ. When $\rho = 1$ these deviations are permanent, y_t follows a random walk. Hence μ is not defined, and variance of y_t tends to infinite. However, if $|\rho| < 1$, the series will return to their mean and the variance of y_t will be finite.

Using the test statistic Γ_μ, the null hypothesis will be accepted if calculated test statistic is greater than the tabulated critical values in Fuller (1976).

In the above cases if the null hypotheses are accepted, then y_t will be a random walk variable without intercept term. Instead, the AR(1) process might include a deterministic trend in addition to the nonzero mean.

- If the maintained model is $y_t = \vartheta_1 + \vartheta_2\, t + \rho\, y_{t-1} + e_t$, then

the null hypothesis $H_0 : \rho = 1$, or $\Delta y_t = \vartheta + e_t \Rightarrow$ a random walk with drift, the first difference of y_t is stationary.
the alternative hypothesis $H_1 : |\rho| < 1$, or $(y_t - \mu - \beta t) = \rho\ (y_{t-1} - \mu - \beta(t-1)\ ) + e_t \Rightarrow$ the deviations of y_t from a linear time trend follow a stationary AR(1) process.

In the maintained model if $|\rho| < 1$ and $\vartheta_2 \neq 0$, then y_t is generated by a trend stationary (TS) process, but if $\rho = 1$ and $\vartheta_2 = 0$, then y_t is generated by a difference stationary (DS) process.

Perman uses the test statistics Γ_Γ, Φ_2, and Φ_3 to test the null of non-stationarity.

Test.1: H_0: $\rho = 1$ is tested using Γ_Γ. The null hypothesis is accepted for the greater calculated test statistic.

Test.2: H_0: $\vartheta_1 = \vartheta_2 = 0$, $\rho = 1 \Rightarrow$ random walk without drift. This is tested using Φ_2, the F statistic. The null hypothesis is accepted for the smaller values of calculated Φ_2 than the critical values given in Dickey and Fuller (1981).

Test.3: H_0: $\vartheta_2 = 0$, $\rho = 1 \Rightarrow$ random walk with drift. This is tested using Φ_3, the F statistic. The null hypothesis is accepted for the smaller values of calculated Φ_3 than the critical values given in Dickey and Fuller (1981).

If the estimated coefficient on time is significant, then y_t contain a deterministic trend. The time series, which contains a time trend should be detrended prior to further modelling.

Nelson and Plosser (1982) discuss trends and random walks in macroeconomic series. In their seminal paper they consider non-stationarity in the mean of series since most of the series show consistent tendency and move away from any initial point as time increases. They constructed the null hypothesis assuming a random walk with drift. Although Dickey and Fuller allow a test of difference stationary hypothesis against trend stationary hypothesis under the assumption that only AR terms are required to obtain satisfactory results, Nelson and Plosser emphasize the use of both hypotheses in a common model. A trend stationary process with first order autoregressive deviations and a random walk with drift can be expressed as

$$z_t = \alpha + \beta t + (1-\phi L)\, u_t \,, \qquad (20)$$

where α and β are fixed parameters, L is the lag operator.

If the trend stationary hypothesis is correct, then $|\phi| < 1$,

If the difference stationary hypothesis is correct, then $|\phi| = 1$, then the model is

$$z_t = z_{t-1} + \beta + u_t \qquad (21)$$

Nelson and Plosser estimate the following model [3]

$$z_t = \rho z_{t-1} + \mu + \gamma t + u_t \quad \text{with } u_t \sim \text{iid } (0, \sigma^2). \qquad (22)$$

The null hypothesis H_0: $\rho = 1$ and $\gamma = 0$ is equivalent to testing for $|\phi| = 1$. Under the null hypothesis the usual t-statistics are not valid. Nelson and Plosser found that standard testing methods lead to finding stationarity around a trend, and they reject the non-stationarity hypothesis in favour of stationarity.

Nelson and Plosser argue that if the series belong to the trend stationary class, then the deviations from trend must be sufficiently autocorrelated to make it difficult to distinguish them from the difference stationary class on the basis of sample autocorrelation. Indeed, the fundamental difference between the two classes of non-stationarity can be illustrated in terms of the roots of the autoregressive moving average (ARMA) polynomials. In that case, if a series is generated from a linear trend stationary process, we should accept the hypothesis of a unit MA root in the ARMA model for its first differences. If time series data are generated by a first order differenced stationary process, we should accept the hypothesis of a unit AR root in the ARMA model for its levels.

[3] The original notations in Nelson and Plosser (1982) have been used.

b. The Augmented Dickey and Fuller (ADF) Tests:

The Dickey and Fuller test is based on the assumption that the disturbance term is white noise. However, if this assumption does not hold, then the Dickey and Fuller regression cannot be used to test for a unit root.

Consider the following regression that is called the Augmented Dickey and Fuller regression

$$\Delta y_t = \alpha + \beta t + \eta y_{t-1} + \theta_1 \Delta y_{t-1} + \theta_2 \Delta y_{t-2} + ..+ e_t \qquad (23)$$

$$= \alpha + \beta t + \eta y_{t-1} + \sum_{i=1}^{q} \theta_i \Delta y_{t-i}$$

where q is chosen to ensure white noise residuals in the regression. Lagged first differences are included in the model to eliminate serial correlation. The lag length can be chosen using Lagrange multiplier test for serial correlation.

Augmented Dickey and Fuller test includes more dynamic than the Dickey and Fuller regression. Dickey and Fuller (1981) demonstrate that the t statistic for $\eta = 0$ from (23) has the same nonstandard limiting distribution if $\rho = 1$. The critical values for the test statistic are identical to the critical values of Γ_Γ.

In equation (23) if the coefficient on time trend is different from zero, then the first differences will be time dependent. This means that the rate of change in y_t is deterministic and it increases with time t. In that case finding unit root will not be sufficient in order to conclude that the first differences are stationary. Nelson and Plosser (1982) explain that a significant time trend occurs very rarely in the presence of a unit root.

Following the steps given by Holden and Perman (1994), the test procedure for unit root testing is:

Step.1: Estimate the equation (23) including sufficient lagged first differences to eliminate serial correlation.

Step.2: Use the test statistic Φ_3 to test H_0: $(\alpha, \beta, \eta) = (\alpha, 0, 0)$ against H_1: $(\alpha, \beta, \eta) \neq (\alpha, 0, 0)$.
If the null hypothesis is rejected, then go to step.3.
If null hypothesis is accepted, then go to step.5.

Step.3: Test $\eta = 0$ using the test statistic from step.1 and compare it with the critical values from standard normal tables.
If the null hypothesis is accepted, then β is nonzero and η is zero.
If the null hypothesis is rejected, then go to step.4.

Step.4: Use the calculated t statistic from step.1 to decide if β is zero, or not.
If the null is accepted, then the y_t is a stationary series without trend.
If the null is rejected, then the y_t is a stationary series with a linear trend.

Step.5: Given an acceptance of the null hypothesis H_0: $(\alpha, \beta, \eta) = (\alpha, 0,0)$ implies that the series has a unit root without trend but with possible drift. Use a t statistic from step.1 to test for $\eta=0$ assuming β is zero and compare it with the nonstandard critical values.

Step.6: Perform a Φ_2 test for H_0: $(\alpha, \beta, \eta) = (0, 0,0)$.
If the null is accepted, then the series is a random walk without drift.
If the null hypothesis is rejected, then the series is a random walk with drift.

Step.7: Impose the zero restriction on β and test H_0: $(\alpha,\eta) =0$ against H_1: $(\alpha, \eta) \neq 0$ by the test statistic Φ_1.
If the null is accepted, then conclude that the series is a random walk without drift.

The calculated Φ statistics are the F tests. If the lagged first differences are included, the calculations of the test statistics are straightforward. Imposing the relevant restrictions, variable deletion tests will give the calculated F values for each specification. However, if no lags of first differences are included, then the calculations of Φ_2 and Φ_1 will be given by the standard F values.

c. The Phillips and Perron (PP) Tests:

Phillips and Perron (1988) use a nonparametric method to correct the serial correlation of the disturbances. The test is based on the estimate of the long run variance of residuals. Their modification of the Dickey and Fuller Γ test is called $Z(\Gamma)$ test. The critical values for Γ_Γ and $Z(\Gamma_\Gamma)$ are the same if the residuals are generated by an independent and identical process.

Although the Phillips and Perron tests and the Dickey and Fuller tests provide identical results, the power of the (Augmented) Dickey and Fuller tests is more than the Phillips and Perron tests at the presence of negative moving average components. However, the (Augmented) Dickey and Fuller tests have low power at the presence of moving average terms. Both the Γ_Γ and $Z(\Gamma_\Gamma)$ statistics should be used to test for a unit root if the diagnostic tests results support the non-normality and serial correlation hypotheses.

d. Park' J Tests:

Park (1990) develops a J(p,q) test using a variable addition method. The test is based on spurious regression results, assuming that a time series has deterministic time polynomials up to the order p being either zero or one, and additional time polynomials are spurious trends.

The J(p,q) test is defined as the ratio of the calculated value of F test to the sample size. The null of the difference stationarity is accepted against the alternative of trend stationarity if the value of J test is large.

The calculation of the Park J tests are easier than the (Augmented) Dickey and Fuller tests and the Phillips and Perron tests since these tests do not require the estimation of the long run variances and the order of lag length.

Tests for the null of stationarity:

a- Park's G test
b- The KPSS test.

These tests assume the null of stationarity, or trend stationarity.

a- Park' G Tests:
Park (1990) G(p,q) tests are based on the same spurious regression results as Park's J tests. However, unlike the J tests, G tests require the estimation of long run variance. This test has asymptotic chi-square distribution with the degree of freedom q-p. Under the alternative hypothesis a series is assumed to be difference stationary after the deterministic trend.

b- The Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) Tests:
KPSS (1992) propose a test of the null hypothesis that series are stationary around a deterministic trend. The test is the one sided lagrange multiplier (LM) test of the null of trend stationarity corresponds to the hypothesis that the random walk has zero variance. Phillips (1987), and Phillips and Perron (1988) initiated the KPSS tests. The asymptotic distribution of the statistic is derived under the null and alternative hypotheses that the series is difference stationary. The critical values for the test statistics are in Table.2, KPSS (1992).

Unit Root Tests with Structural Breaks:

Since the publishing of the seminal paper of **Nelson and Plosser (1982),** it has been argued that most of the macroeconomic series exhibits some kind of stochastic non-stationarity. The total variability of time series is explained by variations in permanent shocks. However, **Perron** emphasizes the presence of a sudden and big change in the mean of time series due to the presence of important structural breaks in the their trend function. Accordingly, he interprets that shocks have no permanent effects, but only the one time shift in trend function is permanent.

Perron (1988, 1989, 1990, and 1994) develops detrending method by considering the implication of structural breaks for the unit root tests. That is, a structural change in trend variable allows a change in the form of level and/or growth rate.

Perron (1988 and 1989) considers the "crash model" and develops a test procedure for a unit root allowing the possibility of one time structural change in trend function. Perron examines the data generating process for y_t in the form that

$$y_t = \alpha y_{t-1} + e_t + \theta e_{t-1} \qquad (32)$$

and concludes that if θ=-1 and α=1, then the process becomes indistinguishable from a stationary process and it is called as "nearly stationary" series. In other words, Perron argues that in finite samples, any trend stationary process is nearly observationally equivalent to a unit root process where disturbances have a moving average component with a root near minus one.

Perron (1990) concentrates on the effects of a changing mean on the tests for a unit root. He finds that the permanent change in the mean of the series might be either instantaneous or gradual. In an instantaneous change, the mean of the series is not affected by the error dynamics; whereas in a gradual change, the mean of the series is affected by the dynamic specification of the error term.

Perron (1990) introduces two models that have different implications with respect to the effects of one time structural change in the trend function. These are

1. **Additive Outlier Model (AOM)**: The change to the new trend function occurs instantaneously.
2. **Innovational Outlier Model (IOM)**: The change to the new trend function occurs gradually.

The AOM is totally different than the IOM with respect to the transition path and the statistical test procedure for a unit root.

Perron (1994) re-examines the implications of structural breaks for the unit root tests considering three kinds of changes in the trend function of a series. These are: a) a change in intercept term; b) a change in slope; c) both.

The idea behind this approach is as the following: the trend function can be determined only by long term economic variables such as population, economic organizations, capital accumulations, technological change, etc. and these variables change very slowly. Under the assumption of one time exogenous change in the trend function, it

would be possible to take the shock out of the disturbance term and put it into the trend function.

Perron's unit root tests do not allow the possibility of one or more structural change under both hypotheses. The test procedures are based on an ARMA process estimating by the ordinary least squares (OLS) and the test statistic is the t statistic. The critical values are presented in Perron (1989) in tables IVB, VI.B, and Perron (1993).

6. Vector Autoregression (VAR) and Vector Error Correction (VEC) Models:

6.1 Vector Autoregression (VAR) Model:

The vector autoregression (VAR) model is used for analyzing the interrelation of time series and the dynamic impacts of random disturbances (or innovations) on the system of variables.

Following Enders (1995), consider a simple bivariate first order VAR, i.e. VAR(1), model

$y_t = \beta_{10} - \beta_{12}x_t + \alpha_{11}y_{t-1} + \alpha_{12}x_{t-1} + u_{yt}$ (36)

$x_t = \beta_{20} - \beta_{21}y_t + \alpha_{21}y_{t-1} + \alpha_{22}x_{t-1} + u_{xt}$ (37)

where it is assumed that both y_t and x_t are stationary; u_{yt} and u_{xt} are white noise with standard deviations of σ_y and σ_x, respectively; u_{yt} and u_{xt} are uncorrelated.

Equations (36) and (37) constitute a two variable first order VAR model. In this system y_t is influenced by current and past values of x_t, and x_t is influenced by current and past values of y_t.

Thus the VAR(1) model captures the feedback effects allowing current and past values of the variables in the system.

The coefficients β_{12} and β_{21} represent the contemporaneous effects of a unit change of x_t on y_t and of y_t on x_t, respectively;
α_{12} is the effect of a unit change of x_{t-1} on y_t,
α_{21} is the effect of a unit change of y_{t-1} on x_t.

Hence y_t and x_t have mutually contemporaneous effects on each other in the system.

The disturbance terms u_{yt} and u_{xt} are shocks or innovations in y_t and x_t. The term u_{yt} has an indirect contemporaneous influence on x_t if $\beta_{21}\neq 0$, and u_{xt} has an indirect contemporaneous effect on y_t if $\beta_{12}\neq 0$.

Equations (36) and (37) represent **the structural VAR model.** This model uses economic theory to describe the dynamic relationship between variables. However, appearance of the endogenous variables on both sides of the equations complicates the estimation and inference processes. A standard VAR model can be applied to overcome the difficulties of the structural VAR model.

The standard form of VAR model for the two variable case can be written as

$$y_t = \gamma_{10} + \gamma_{11} y_{t-1} + \gamma_{12} x_{t-1} + \varepsilon_{1t} \quad (38)$$
$$x_t = \gamma_{20} + \gamma_{21} y_{t-1} + \gamma_{22} x_{t-1} + \varepsilon_{2t} \quad (39)$$

In (38) and (39), the terms ε_{1t} and ε_{2t} are random innovations or shocks, and they are correlated if there are contemporaneous effects of y_t on x_t and of x_t on y_t, but the terms ε_{1t} and ε_{2t} are uncorrelated if there are not contemporaneous effects on each other.

In the system each endogenous variable is determined by a function of the lagged values of the two endogenous variables. The OLS is the appropriate method since only lagged variables are included on the right hand side of the each equation, and also disturbances are assumed to be serially uncorrelated with constant variance.

Two questions arise about the construction of a general VAR model.
First, how can we determine the set of variables to include in a VAR model?
Second, how can we determine the appropriate lag length?

The included variables in a VAR model are selected according to the relevant economic theory. The selected variables must have economic influences on each other. In other terms, there must be causality between them.

The overparameterization and loss of degrees of freedom problems must be avoided to capture the important information in the system.

The appropriate lag length must be determined by allowing a different lag length for each equation at each time and choosing the model with the lowest AIC and SBC values. The same sample period must be considered for different lag lengths. If the lag length is too small, the model will be misspecified; if it is too large, the degrees of freedom will be lost.

The VAR analysis determines the interrelationship among the economic time series rather than the parameter estimates.

The residual correlation in the VAR model reveals the interaction of the variables in the previous periods.

The main uses of the VAR model are **the impulse response analysis, variance decomposition, and Granger causality tests.**

An impulse response function traces the response of the endogenous variables to one standard deviation shock or change to one of the disturbance terms in the system.

A shock to a variable is transmitted to all of the endogenous variables through the dynamic structure of the VAR.

Therefore, an impulse response function shows the interaction between/among the endogenous variables sequence.

Variance decomposition analysis provides information about the dynamic behaviour of the model and the relative importance of each random disturbances or innovation in the VAR.

Variance decomposition shows the proportion of the movements in the endogenous variable sequence as a result of its own shocks against shocks to other variables.

VAR models are used to test the causality relationship between the variables in the system.

Granger causality provides important information about the exogeneity, in other words x_t is defined as an exogenous variable if the current and past values of y_t do not affect x_t.

In that case, all the coefficients on current and past y_t are zero.

Granger noncausality shows that x_t sequence is independent of both the u_{yt} shocks and y_t sequence.

6.2 Vector Error Correction (VEC) Model:

Engle and Granger (1987) point out that a linear combination of two or more nonstationary series may be stationary. The stationary combination may be interpreted as the cointegration, or equilibrium relationship between the variables.

For example, reconsidering the consumption model in the previous section, if the consumption and income are cointegrated, then there exists a long run relationship between them.
However, if they are not cointegrated, then consumption might drift above or below income in the long run, implying that consumers either spend too much or increase savings.

A VEC model is a restricted VAR model. The VEC specification restricts the long run behaviour of the endogenous variables to converge to their long run equilibrium relationships and allow the short run dynamics.

Consider the relationship between consumption and income in a simple EC model

$\Delta C_t = \theta_1(C_{t-1} - \lambda Y_{t-1}) + u_{1t}, \theta_1>0$ (40)

$\Delta Y_t = -\theta_1(C_{t-1} - \lambda Y_{t-1}) + u_{2t}, \theta_2>0$ (41)

where u_{1t} and u_{2t} are white noise disturbances, θ_1 and θ_2 represent the speed of adjustment parameters. θ_1, θ_2 and λ are the positive parameters.

The cointegrating term $(C_{t-1} - \lambda Y_{t-1})$ is the error correction term since the deviation from long run equilibrium is corrected gradually through short run adjustments. C_t and Y_t are the two endogenous variables.

In an EC model, **the short run dynamics** of the variables in a system are influenced by the deviations from the long run equilibrium. For example, C_t and Y_t change in response to the previous period's deviation from long run equilibrium.

In the VEC model if:

- the deviations are positive, i.e.$(C_{t-1} - \lambda Y_{t-1}) >0$, then the level of income would rise and the level of consumption would fall, as the other things are constant. Long run equilibrium is achieved as $C_{t-1} = \lambda Y_{t-1}$.
- $C_{t-1} = \lambda Y_{t-1}$, then C_t and Y_t change only in response to u_{1t} and u_{2t} shocks.
- θ_1 is large, then C_t shows greater response to the previous period's deviation from long run equilibrium.
- θ_1 is small, then C_t is unresponsive to the previous period's deviations from equilibrium.
- $\theta_2 =0$, then Y_t changes only in response to u_{2t}, since $\Delta Y_t=u_{2t}$. Hence C_t changes to eliminate any deviations from long run equilibrium.
- $\theta_1 =0$ or $\theta_2 =0$, there would not be a causality relationship between cointegrating variables.
- $\theta_1 =0$ and $\theta_2 =0$, there would not be a long run equilibrium relationship between the two variables. The VEC or cointegration models cannot be used for these variables.

The crucial point of using VEC models is the requirement of cointegration between the two variables with the cointegrating vector (1 -λ). In other words, (C_{t-1} - λY_{t-1}) must be stationary.