# DATA HANDLING AND TIME SERIES MODELLING


Presented By:

Oduh,Moses Onyema

*DEPARTMENT OF ECONOMICS,*

*UNIVERSITY OF NIGERIA*

*&*

*AFRICAN INSTITUTE FOR APPLIED ECONOMICS,ENUGU*

# DATA Analysis

- Time series
  - A time series is a set of observations on the values that a variable takes at different times e.g. daily (stock prices), weekly (money supply), monthly (unemployment, CPI),quarterly(GDP), annually (budgets), quinquennially (i.e 5 years manufacturers) or decennially (census)

# Data Analysis….contd

- Cross-Section Data
  - These are data on one or more variables collected at the same point in time, such as census of population.
  - The major distinguishing characteristics between time series and cross-section is that in time series same variable assumes different values at different point in time, while cross-section variable assumes different value at the same point in time.

# Data Analysis…..contd

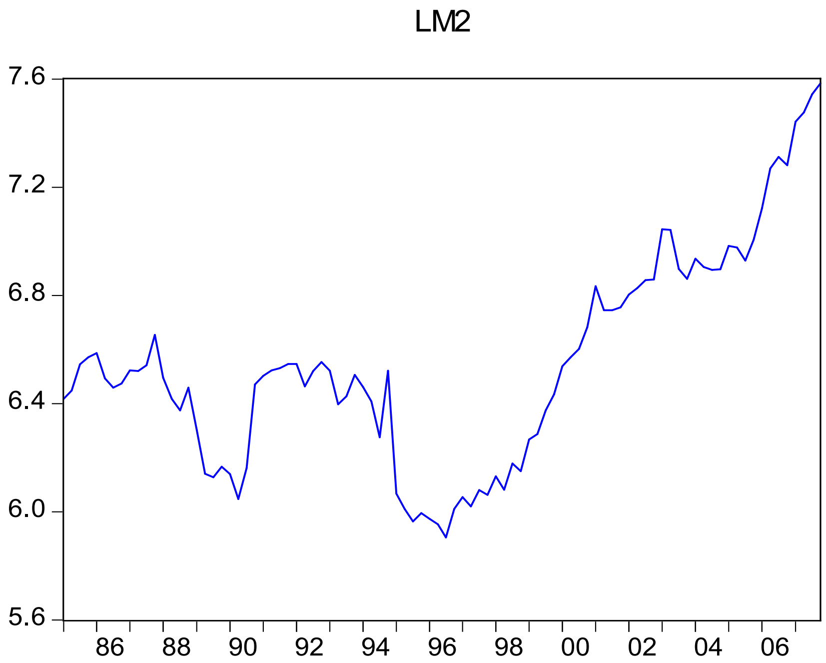
- Pooled Data
  - Pooled or combined data is the combination of time series and cross-section. That is, they are elements of time series and cross section data.

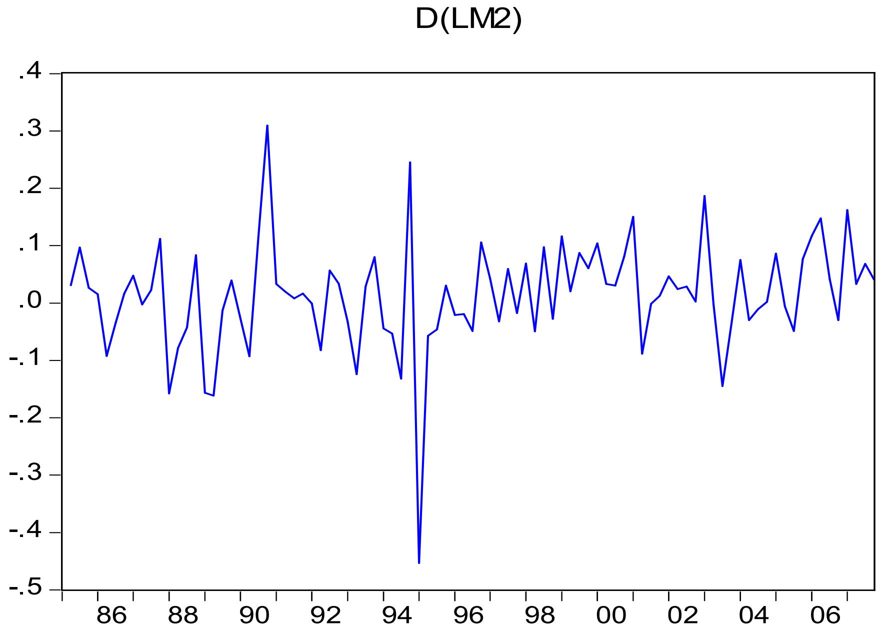
# Data Analysis.....contd

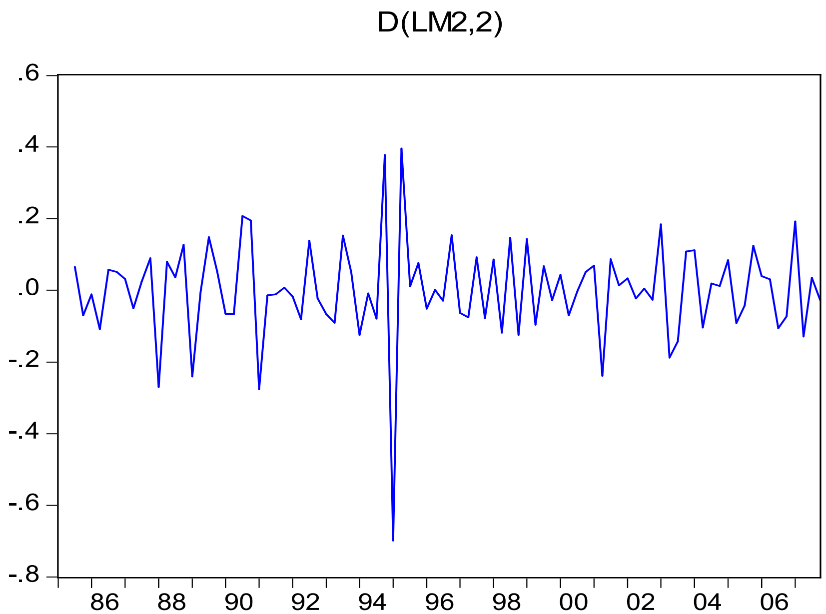
- **Data screening**
- **Time series characteristics and data behaviour**
  - Data stationarity
  - Data co integration
  - Long run economic relations
  - Detrending/deseasonalising

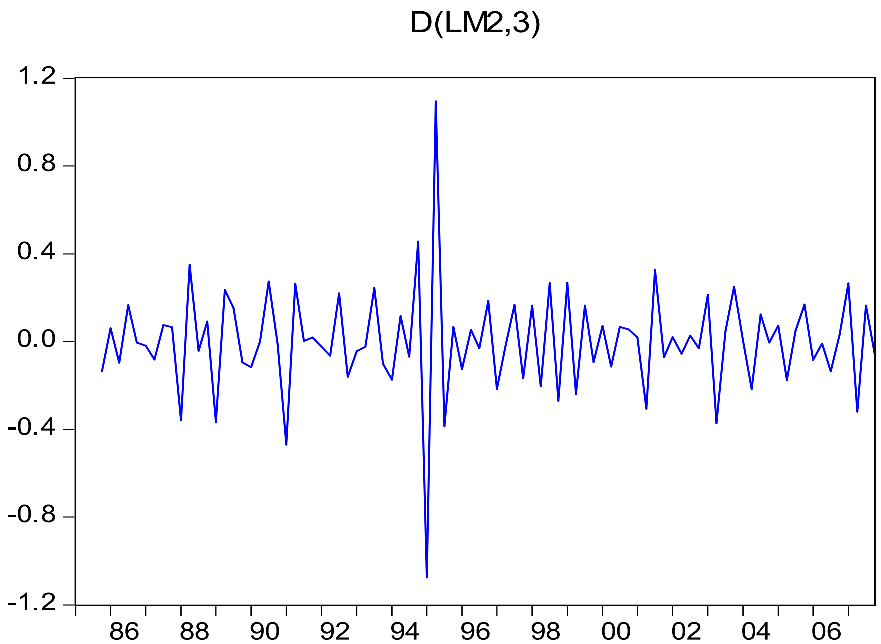
# Stationary of data

- Defining stationary data
  - A *stationary* time series is one whose statistical properties such as mean, variance, autocorrelation, etc. are all constant over time. Most statistical forecasting methods are based on the assumption that the time series can be rendered approximately stationary
  - The value of variance between the two time periods depends only on the distance or gap or lag between the two time periods and not the actual time at which the covariance is computed
  - There is no influence of time on economic variables that contributed to their having different value at different point in time


LM2
7.6
7.2
6.8
6.4
6.0
5.6
86 88 90 92 94 96 98 00 02 04 06


D(LM2)
.4
.3
.2
.1
.0
-.1
-.2
-.3
-.4
-.5
86 88 90 92 94 96 98 00 02 04 06


D(LM2,2)
.6
.4
.2
.0
-.2
-.4
-.6
-.8
86 88 90 92 94 96 98 00 02 04 06


D(LM2,3)
1.2
0.8
0.4
0.0
-0.4
-0.8
-1.2
86 88 90 92 94 96 98 00 02 04 06

# Why Stationary data ?

- To be able to obtain meaningful sample statistics such as means, variances, and correlations with other variables. Such statistics are useful as descriptors of future behavior *only* if the series is stationary.
- To be able to have a consistent and reliable forecast of economic variables
- To have a stable and efficient prediction on the movement of economic variables

# Problems of non-stationary

- **Spurious or nonsense regression**
- **Problem of autocorrelation**
- **Problem of random work – if the best prediction of a variable  tomorrow is equal to its value today plus a purely random shock (error term). That is, forecast of such variable would be a futile exercise.**
- **Inconsistent ,unreliable and unstable forecast of economic parameters**
- **Non-mean reverting(or reversing)**

# Possible reasons for non-stationary

- Institutional factors
  - Data handling and error of entry
  - Poor data screening
  - Missing data and data improvise
  - Data mining and fishing
  - Merging data from different sources
- Political regime handover and takeover
  - macroeconomic policy inconsistency
  - Data manipulation for political legitimacy

# Dealing with non-stationary data

- There are two ways of dealing with non-stationary series:
  - Differencing; and
  - Detrending
- Non-stationary data is made stationary by data differencing
- Differencing is the process of removing the influence of time from economic variables
- Detrending is the process of removing the influence of time and seasonal differences by incorporating time parameters in econometric models

# The **BIG** debate about differencing and detrending

- Differencing, a data mining and fishing?
- Lost of original data and behavioral pattern of the data series?
- Detrending, trend stationary (TS) Vs difference stationary (DS)
- Detrending, stochastic Vs deterministic

# How do we know non-stationary data?

- Dickey-Fuller (DF) and Augmented Dickey-Fuller (ADF) – unit root test
- Graphical Analysis
- Autocorrelation Function (ACF)
- Trend analysis

# Unit Root (DF and ADF)

- A series is said to have a unit root if that series is non-stationary
- When a series is stationary we say that the series is integrated (stationary of order zero)
- The process of making a non-stationary series to be stationary is by differencing
- The number of times a series is differenced before it became stationary is the series order of integration
- If a series is differenced once we say that the series is integrated of order (1); two times difference is integrated of order (2) and “d” times difference is said to be integrated of order “d”.

# DF AND ADF Unit Root Test

$$\mathrm{Y} = \alpha \mathrm{Y}_{\text{t-1}} + \mu \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

$$\mathrm{Y} = \beta + \alpha \mathrm{Y}_{\text{t-1}} + \varepsilon \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

$$\mathrm{Y} = \beta_1 + \alpha_1 \mathrm{Y}_{\text{t-1}} + \alpha_2 \mathrm{T} + \varepsilon \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

$$\mathrm{Y} = \alpha \mathrm{Y}_{\text{t-1}} + \sum_{i=1}^{p} \beta \Delta Y_{t-i} + \mu \ldots\ldots (4)$$

# Hypothesis of Unit root

- Tau statistics (T)
  - If the t calculated of the data series is greater than the (DF) tau statistics reject the non-stationary hypothesis (Ho)
  - Otherwise the series is stationary
  - Intuitive reasoning of unit root
    - If the parameter is statistically (1 or unit) then:
      - » Y =1(Yt-1) +u =ΔYt=u (this is a random work signifying that the series is not stationary)
      - » Y =0(Yt-1) +u =Yt=u

# Difference stationary VS Trend stationary

- $Y=a+bY_{t-1}+cT+u$

where **"T"** is trend;**"c"** is the trend parameter

1)If a series has a trend whose parameter is significant in a statistical sense we say that the series is trend stationary (TS) meaning that the model can only be estimated by including trend- detrending (removing the influence of time – stationary) – trend is deterministic

2) If a series has a trend whose parameter is **NOT** significant in a statistical sense we say that the series is **DIFFERENCE** stationary (DS) meaning that the model can only be estimated by differencing the trend- (removing the influence of time – stationary) – trend is stochastic

# Regression with stationary series

- When we have a series that are integrated of order(d); where "d"$\geq$0 eg: $Y=aX+bZ+cQ+dR+u$.......(4)
- where, $Y\sim I(1)$;$X\sim I(1)$; $Z\sim I(1)$,$Q\sim I(2)$;$R\sim I(0)$
- Estimate:

$\Delta Y = a\Delta X+b\Delta Z+c\Delta\Delta Q+dR+e$......(5)

Is there anything wrong with estimation of equation (5) ? **"YES and NO"**

# THE BIG YES – Co-integrated series

- It is possible to have a situation where the series , of course with the same order of integration are not stationary individually but their linear combination is stationary
- If the scenario above arises then we say that the series whose linear combination is stationary are co-integrated
- When co-integration exist then, the series are tied together in the long run – long run relation

# THE BIG NO – "Co-incidental" series

- The series have the same order of integration by chance
- They do not have any relationship or tied together in the long run
- Estimation of (5) is well intended and articulated
- Series are estimated in their order of integration – no linear combination

# Linear combination and co-integration

- Linear combination and test of co-integration
  - Granger residual test
    - Identify all the series with the same order of integration
      - **[Y –X-Z]=u……(6)**
- "u" is the linear combination of all the series with the same order of integration
- u=λut-1…..(7)
- If equation (7) is stationary of integrated the, the linear combination (6) is integrated and is said to be co-integrated.

# Co-integration and Error Correction Model (ECM)

- If series are co-integrated then, there is need to untie the series in the long run
- This is done by estimating an Error Correction Model (ECM)
- Incorporation the linear combination of the error term from the linear combination.

# Error Correction Model (ECM)

- Granger two-step procedure
  - First run an OLS regression of all the variables integrated of the same order
  - Save the residual from the regression above
  - Incorporate the first lag residual in the final equation
  - The parameter of linear combination is the ECM parameter
  - The parameter shows the magnitude of error corrected every period
  - The parameter is usually negative

# Steps in ECM estimation

- Estimate $u = \lambda u_{t-1}$ ---------(7)
- $\Delta Y = a\Delta X + b\Delta Z + c\Delta\Delta Q + dR$ **$-\theta u_{t-1}$** $+ \varepsilon$..(8)
- Equation (8) is the error correction model while **$\theta u_{t-1}$ is the error correction mechanism.**

# Conclusion

- Order of integration is a necessary condition for co-integration , but not sufficient condition
- Co-integration is a necessary and sufficient condition for ECM
- More importantly, lags in the model is dependent on the choice of model and optimal lag length of the choice model