Overview of financial econometrics

Presentation at the Faculty Development Programme on Financial Econometrics held at
Bapuji Institute of Engineering & Technology (BIET), Davangere, Karnataka, India, on December 9-10, 2015.

Stephen ARO-GORDON, Ph.D.
Baze University Abuja, Nigeria
Department of Financial Mathematics
Email: stephen.aro-gordon@bazeuniversity.edu.ng / getyouthsworking@yahoo.com

WORKSHOP NOTES

BAZE UNIVERSITY ABUJA

December 9-10, 2015 @ BIET, Davangere, India

Objectives…

BAZE UNIVERSITY ABUJA

1. To provide an overview of financial econometrics / time series analytical models with emphasis on Granger causality analytics.
2. To experiment with the use of appropriate econometric software (*EViews*) in performing uncomplicated time series analysis, focusing on single time series data.

Anticipated learning experience…

BAZE UNIVERSITY ABUJA

1. At the end of this workshop, you should have appreciable knowledge of time series analytical models with emphasis on Granger causality analytics.
2. You should also be able to use an appropriate econometric software, notably the *EViews*, to perform simple time series analysis.

Coverage…

BAZE UNIVERSITY ABUJA

Session	*Activity*
I	Conceptual reviews - introduction to financial econometrics
II	Practical session – dealing with selected cases and problems using econometric software

Key application areas…

1. Asset returns
2. Asset pricing – technical analysis & fundamental analysis
3. Portfolio spreads
4. Interest rates / yields
5. Yield of a portfolio of bonds
6. GDP and other economic series, etc.

Session 1

Introduction to financial econometrics

- Statistics primer

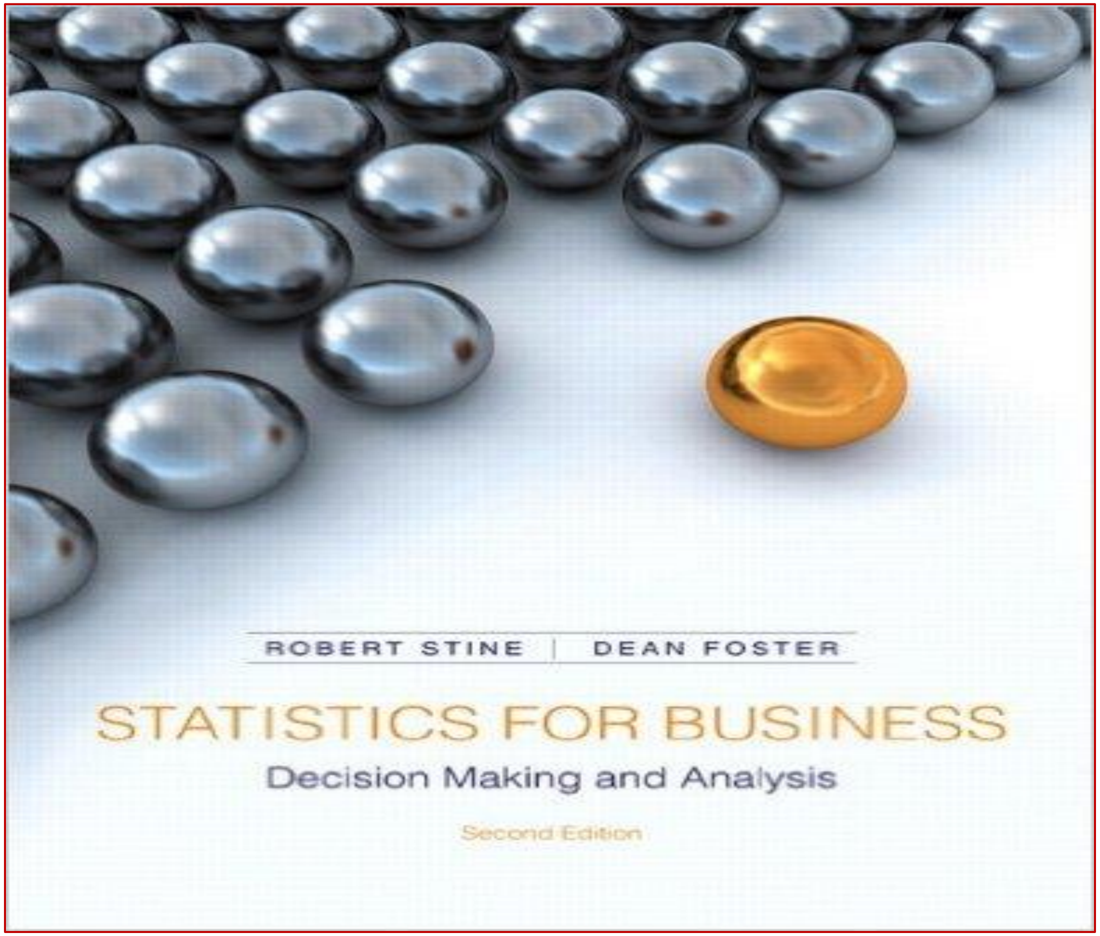
BAZE UNIVERSITY ABUJA

Primary references

Useful texts & references

Resources...

BAZE UNIVERSITY ABUJA



Resources

BAZE UNIVERSITY ABUJA

- *Introductory Econometrics for Finance, 2nd edn.* by Chris Brooks (2008). Cambridge University Press.
- *Statistics for Business: Decision-making and analysis* by Robert Stine & Dean Foster (2012).
- *Business Statistics*, by S. L. Aggarwal & S. L. Bhardwaj (2010). New Delhi: Kalyani Publishers

BAZE UNIVERSITY ABUJA

Great inputs from:

The Workshop on Financial Econometrics, Correlation, Causation and Co-integration
@SDMIMD, Mysore, India on 20th August 2015
delivered by

Dr Kuldeep Kumar
Professor, Department of Economics & Statistics,
Bond University, Australia. Email: kkumar@bond.edu.au

December 9-10, 2015 @ BIET, Davangere, India

Econometric software: A helpful resource

BAZE UNIVERSITY ABUJA

- ***Introduction to EViews 6.0*** by Anders Thomsen *et al* (January 2013), Analytics Group, Aarhus University, Denmark.
- In comparison with the traditional spreadsheet package that you are familiar with, the ***EViews*** allows you to do more advanced calculations, regressions and simulations
- Relatively easy to use
- Faster calculation time
- So, explore & explore!

On the underlying maths…

BAZE UNIVERSITY ABUJA

- ***Observe:*** We have reduced the mathematical aspects to the barest minimum.
  - We are focusing on practical principles and applications of financial econometrics.
  -However, workshop participants are advised to consult the primary references for more mathematical and technical details.

What can you say about the contributions of these two people?

BAZE UNIVERSITY ABUJA

- Pawel Ciompa
- Ragnar Frisch

What is 'Econometrics'?

BAZE UNIVERSITY ABUJA

?

Statistical methods / econometrics:
… a major field in Financial Mathematics

BAZE UNIVERSITY ABUJA

Classification code	Mathematical finance
91G10	Portfolio theory
91G20	Derivative securities
91G30	Interest rates (stochastic models)
91G40	Credit risk
91G50	Corporate finance
91G60	Numerical methods (including Monte Carlo methods)
91G70	**Statistical methods, econometrics**
91G80	Financial applications of other theories (stochastic control, calculus of variations, PDE, SPDE, dynamical systems)
91G99	None of the above, but in this section

Source: http://www.ams.org/msc/msc2010.html?t=91Gxx&btn=Current [23/11/2015]

What 'Econometrics' is all about…

BAZE UNIVERSITY ABUJA

Some useful definitions:

1. = application of mathematics, statistical methods, and computer science to economic data
2. = a branch of economics that aims to give empirical content to economic relationships, e.g.

- ***Unemployment & economic growth***
- ***Wages & years of education***
- ***Income & Spending***

Basic econometric methods / estimators / tools:

1. ***Linear regression model*** – OLS (ordinary least squares) based on Gauss-Markov assumptions
2. When 'normality' assumption is violated, other estimation techniques are applied:

- Maximum likelihood estimation
- Generalized methods of moments
- Generalized OLS
- Bayesian statistics

What 'Econometrics' is all about…

BAZE UNIVERSITY ABUJA

1. = Quantitative analysis of actual economic phenomena on the basis of theory, observation, and appropriate methods of inference.
2. =Translating data into models to make forecasts and to support decision-making
3. = Sifting through massive data to extract simple relationships

What about Financial Econometrics?

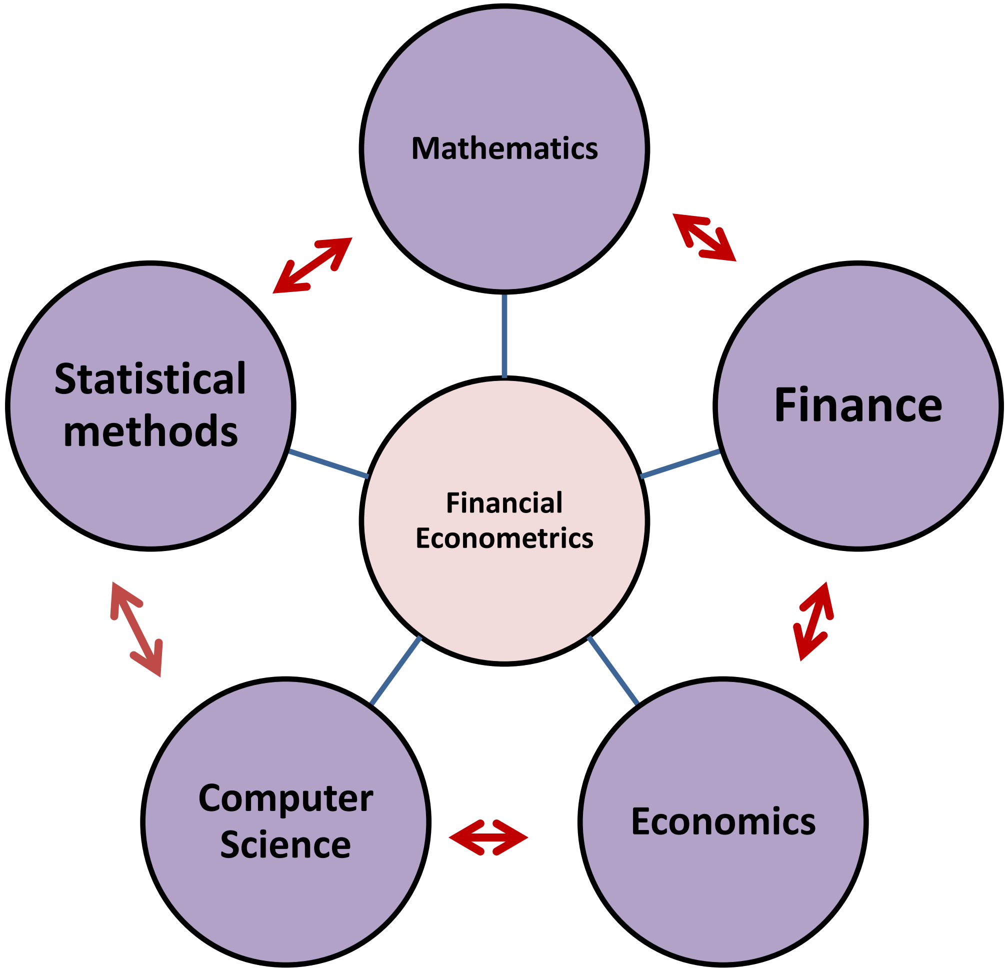
- ***Financial econometrics*** applies mathematical, statistical, and computer science techniques to problems in finance.
- ***Financial econometrics*** studies how the supply of ***capital*** and its use are considered and measured.
- **Note**: ***'Capital'*** is economic resource – a resource or resources that can be used to generate economic wealth (Encarta Dictionaries, 2009).

Common issues in Financial Econometrics?

- Asset valuation – real estate, stocks, bonds, derivatives, currencies and other financial assets – CAPM, APT, EMH, etc.
- Corporate Finance
- Tests of random walk hypothesis
- Term structure of interest rates
- Causality analysis
- Event analysis, etc.

*****Financial Econometrics…

BAZE UNIVERSITY ABUJA



Why we need to follow the protocol in Econometrics

BAZE UNIVERSITY ABUJA

- Observational data versus controlled experiments
  - Systems analysis and control theory
  - Observational study design is not peculiar to econometrics; other disciplines also use the approach
  - The approach basically allows the analyst to do model estimation and investigate a model's empirical consequences.
  - Hence, in econometrics, we use the regression methods because we cannot use controlled experiments.
  - ***Most importantly, data analysis on the basis of observational data should be guided by the study protocol / procedure / - taking models through statistical trials***

... Utility

BAZE UNIVERSITY ABUJA

1. Economics
2. Finance
3. Marketing
4. OB & HRM
5. OM
6. Machine performance
7. Engineering
8. Data analysis
9. *Climate change* – e.g. verifying causal relationship between greenhouse-gas emissions and higher temperatures

(*The Economist*, November 28th-December 4th 2015, "Clear thinking on climate change", p. 10)

… a wide range of job opportunities for analysts

BAZE UNIVERSITY ABUJA

1. Role as Quantitative analysts in public and private organizations
2. Data analysis, structuring and transaction advisory
3. Credit analysis / scoring /provisioning
4. Investment banking/ Corporate finance / trading
5. Asset management / portfolio optimization / trading strategy development
6. Credit cards
7. Risk management
8. Mortgage banks
9. Management Consulting
10. Derivatives pricing and hedging
11. Business/asset valuation
12. Venture capital
13. Foreign exchange services, etc.
14. Operations management
15. Research / academia

Still on job opportunities for analysts…

BAZE UNIVERSITY ABUJA

- Historical analysis of an organization
- Projecting an organization's financial performance
- Project finance
- Real estate
- Oil and Gas projects
- Banking & Financial Institutions
- Personal finances
- Non-profit organizations / NGOs
- Government – at Federal/central/national, State/Regional, local council
- Investment Banking
- Academia - research and educational centres

The essence of financial econometrics - The BIG PICTURE

BAZE UNIVERSITY ABUJA

- Well-being of the society…
- The current massive data availability – the Internet age / smart phones
- Rapid and monumental changes and implications on the future of the professions – e.g. technology 'disruptions' rendering traditional approaches antiquated, opaque, and unaffordable (Susskind &Susskind, 2015)

Financial econometrics: Wide variety of data sets…

BAZE UNIVERSITY ABUJA

- Fundamental economics
- Real estate
- Human resources management
- Accounting
- Advertising
- Agriculture
- Banking & Finance
- Business

- Finance & investments
- Marketing
- Opinion polls
- Transportation
- Sports
- Life sciences
- Physics and engineering, etc.

BAZE UNIVERSITY ABUJA

Errors in data analytics: Areas to watch out for

Minimizing statistical errors:

First, let's recall what statistics is all about

BAZE UNIVERSITY ABUJA

- Statistics is the scientific method of collection, classification, representation, analysis and interpretation of numerical data with a view to making meaningful inferences on the objects on which the data were collected.
- Statistics answers questions using data – not numbers only; we also use pictures, graphs, tables, etc.
- Statistical models are used to simplify reality and help us to answer questions.

- Capacity to make desirable, informed, result-oriented decision.
- one thing is clear: we use statistical processes to serve as guide towards making qualitative decisions.

Are 'numbers' the end?

BAZE UNIVERSITY ABUJA

- "Far too many reviews are dominated by dry discussions of numbers… The review should be a creative exercise, not a drill where people regurgitate data."

- L. Bossidy & R. Charan (2002) in ***Execution: The discipline of getting things done.*** **New York: Crown Business, p. 208**

Minimizing statistical errors Continuation

- ***Wrong methodology*** – A good number of unintentional mistakes come from using the wrong method to address the question at hand thereby leading to misinterpretation of results.
- Choose and use correct statistical methods for every problem

- ***Data organization***: Avoid error here by clearly defining your variables (observations) before you record your data – what does *X1*, *X2*, *X3*, … mean?
- The name of a variable should describe its attributes.

Why exercising good judgement is imperative…

BAZE UNIVERSITY ABUJA

- Analytical tools are as important to the modern executive as pliers and screwdrivers are necessary to the auto mechanic.
- Like a mechanic, the analyst must know his business well enough to choose the proper statistical tool to solve the problem at hand.

What statistical tools would be useful to you for investigating these questions?

- ❑ What are your competitors doing to increase market share?
- ❑ What are your competitors doing differently in the area of customer service?
- ❑ How are your best-in-class competitors handling cost, quality, technology and HR?
- ❑ Does your organization have the cost structure (or capital structure) that will allow it to compete profitably?
- ❑ Assume your organization needs a new organogram; what new sales management skills will be required?

Minimizing statistical errors
Further tips…

BAZE UNIVERSITY ABUJA

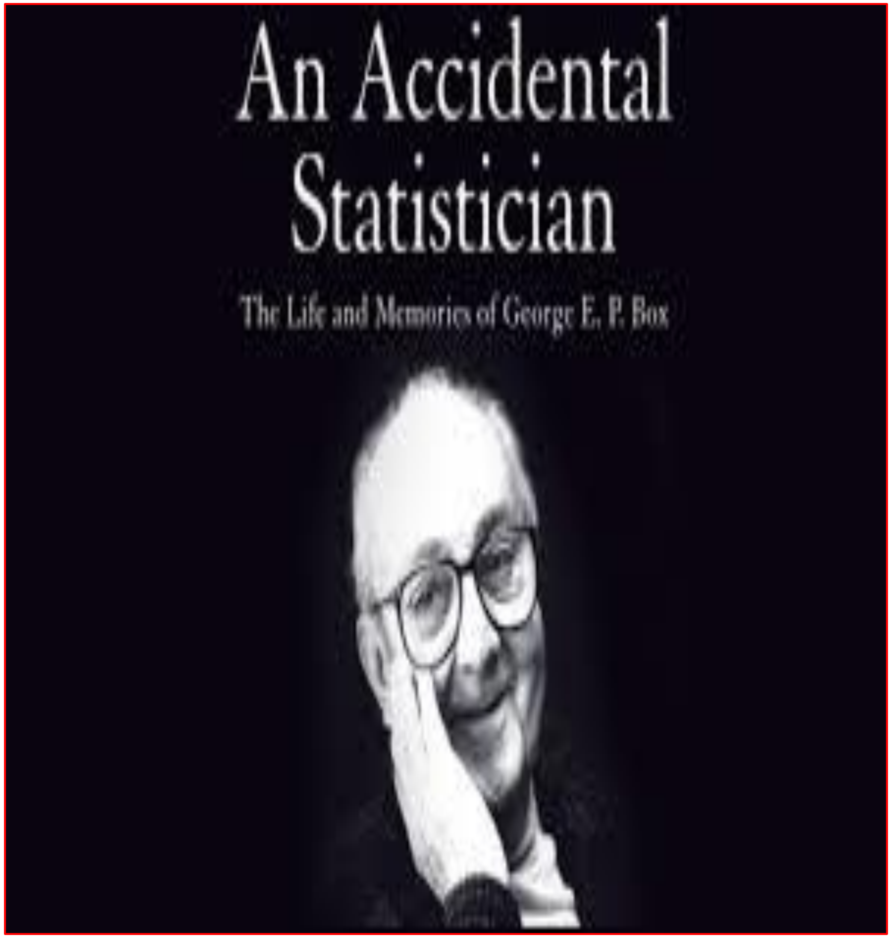
- Understand the underlying business theory / issues / questions first before performing statistical analysis – this is pivotal to drafting meaningful questionnaires.
- The central point of statistics is problem-solving – how are your analyses helping your organization or country to make better decisions or policies?
- Don't carelessly round up data – check to ensure that your data add up – data credibility
- The need to produce better decisions and insights from the massive data amount generated in today's world of business and science.
- Technology – Computers now perform most of the calculations that once dominated statistics and related courses
- Use Statistical packages / software carefully – interpretation of results must make sense to you first before it can make sense to your audience

Minimizing statistical errors

Further tips…

"All models are wrong, but some are useful."

- George E. P. Best



Data normalization and standardization

BAZE UNIVERSITY ABUJA

- Basically, to normalize data, traditionally this means to fit the data within unity (1), so all data values will take on a value of 0 to 1 (***Ben Etzkom, 2012***). Where applicable, data should be normalized or standardized to bring all of the variables into proportion with one another.
- This is important so that the coefficients associated with each variable will scale appropriately to <u>adjust for the disparity in the variable sizes,</u> thereby reflecting meaningful relative activity between each variable, i.e., a positive coefficient will mean that the variable acts positively towards the objective function, and vice versa.

On data normalization and standardization

<u>Emergence of Post-Modern Portfolio Theory (PMPT)</u>

BAZE UNIVERSITY ABUJA

- Associated with software entrepreneurs <u>Brian</u> M. Rom & <u>Kathleen</u> Ferguson, PMPT is an expanded risk-return paradigm designed to address the <u>major practical limitations</u> of CAPM/MPT – the assumption that of a <u>discrete</u>, normal (mean-variance) distribution that may not accurately reflect <u>investment reality</u>.
- Thus, the <u>lognormal</u> distribution was introduced as a more robust model for the pattern of investment returns.

Avoid data confusion / mishandling
Understand data classification / typology…

BAZE UNIVERSITY ABUJA

- ***Qualitative data*** – Categorical - Discrete - Nominal – Ordinal (Likert scale) – we cannot easily measure or count; e.g. gender, behaviour, quality…
- Performing purely quantitative techniques such multiplication and division on categorical data will yield meaningless results.
- Don't put ordinal data in a pie chart!
- Don't carelessly round off data particularly in pie charts.
- Be careful with elaborate graphs
- Clarity in knowledge of Mean, Mode, & Median…

- ***Quantitative data*** – data that we can easily measure and count; e.g. age, weight, height, sales, production output, prices…Numerical – Continuous – Interval – Ratio
- ***Time series (trend analysis)*** – changing values of a variable over time / at different times.
- ***Cross-sectional data*** – data that measure attributes of different objects at the same time – one-shot data.
- ***Panel Data*** – Data collected on various objects (individuals, countries, etc.) for sequential periods – a combination of time-series and cross-sectional data.

Self-study…

BAZE UNIVERSITY ABUJA

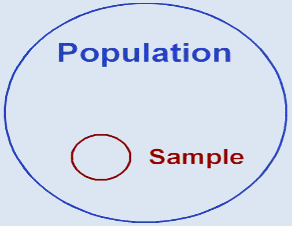
Parametric

versus

non-parametric

- ❑In what situations should you use chi-square?

Ensuring robust sampling



BAZE UNIVERSITY ABUJA

- ❑ The essence of sampling is to obtain maximum, accurate, and reliable information about the universe with the minimum sacrifice of money, time, and energy.
- ➢ Statistical regularity - randomness
- ➢ Inertia of large numbers – more inert (constant) than small ones
- ➢ Persistence – same attributes as the universe
- ➢ Optimization – cost-effective & efficiency
- ➢ Validity – selected at random, scientifically done

Dealing with sampling errors…

BAZE UNIVERSITY ABUJA

- ***Biased errors*** – prejudice of the analyst / investigator
- ***Note:*** Increasing sample size will not cure biased errors.

- ***Unbiased errors*** – accidental or arising in the course of events or survey
- Also called random sampling error – only a part of the universe is ultimately observed.
- ***Solution:*** Ensure that the sample size is reasonably large to neutralize this type of error

Types of statistics revisited …

BAZE UNIVERSITY ABUJA

Types of Statistics

- Descriptive statistics
- Inferential Statistics

Understanding data typology…

BAZE UNIVERSITY ABUJA

- Types of Data
  - Qualitative data
  - Quantitative data
  - Econometric data

Still on data typology…

BAZE
UNIVERSITY
ABUJA

Types of Data

- Primary data
- Secondary data

Revisiting sources of secondary data

BAZE UNIVERSITY ABUJA

S/NO	SOURCES	Examples of data
1	ARCHIVES	General Information
2	PRINT MEDIA (Newspapers, daily, weekly or monthly magazine)	Election result, Academic Result, admission list of Schools, and other news items
3	ELECTRONIC MEDIA	Census result, election result, Radio, television, internet/websites, News items and general information.
4	MUSEUM (a cultural archive)	Information on antiquities, cultural information.
5	**NATIONAL BUREAUS OF STATISTICS PLANNING COUNCILS, CENTRAL BANKS, WORLD BANK, IMF, STOCK EXCHANGES, ETC.**	**Wages and salaries of workers, facts on national economy, asset returns**

Secondary data:

Have they passed the test of scrutiny before usage?

BAZE UNIVERSITY ABUJA

- A major feature of secondary data is that it may be fraught with bias, and may not be so accurate or adequate, depending on the purpose of the investigation.
- Consequently, since the researcher is typically not the original compiler of the data, there is a need for proper editing and scrutiny of the secondary data in order to make it appropriate for usage.
- To do this, answers must be provided to the pertinent questions listed here – (right side of this slide).
- When answers are satisfactorily provided for each of these afore mentioned questions satisfactorily; the secondary data are said to have been *scrutinized* and edited, thus made reliable for use by researchers.

1. What is the type and purpose of the institution which the data emanated?
2. Are the data accurate and adequate?
3. Are the data biased?
4. In what types of units are the data expressed?
5. Are the data related to the problem under study?

What is ‘normal distribution’?

BAZE UNIVERSITY ABUJA

?

Normal Distribution

BAZE UNIVERSITY ABUJA

- Normal Distribution is sometimes called the *Gaussian distribution*.

The density function for this distribution is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)}{2\sigma^2}} \qquad -\infty < x < \infty \qquad (1)$$

where μ and σ are the mean and standard deviation, respectively. The corresponding distribution function is given by

$$F(x) = P(X \le x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(v-\mu)^2}{2\sigma^2}} dv \qquad (2)$$

- We say that the random variable *X* is *normally distributed* with mean μ and variance σ^2 if *X* has the distribution function in the equation (2).

Let *Z* be the random variable corresponding to the following

$$Z = \frac{X-\mu}{\sigma} \qquad (3)$$

- then *Z* is called the *standard normal variable* corresponding to *X*. The mean or expected value of *Z* is 0 and the standard deviation is 1. In such cases the density function for *Z* can be obtained from the definition of a normal distribution by allowing μ = 0 and σ^2 = 1, yielding

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad (4)$$

- This is often referred to as the *standard normal density function*.

The corresponding distribution function is given by

$$F(z) = P(Z \le z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{v^2}{2}} dv = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_{0}^{z} e^{-\frac{v^2}{2}} dv \qquad (5)$$

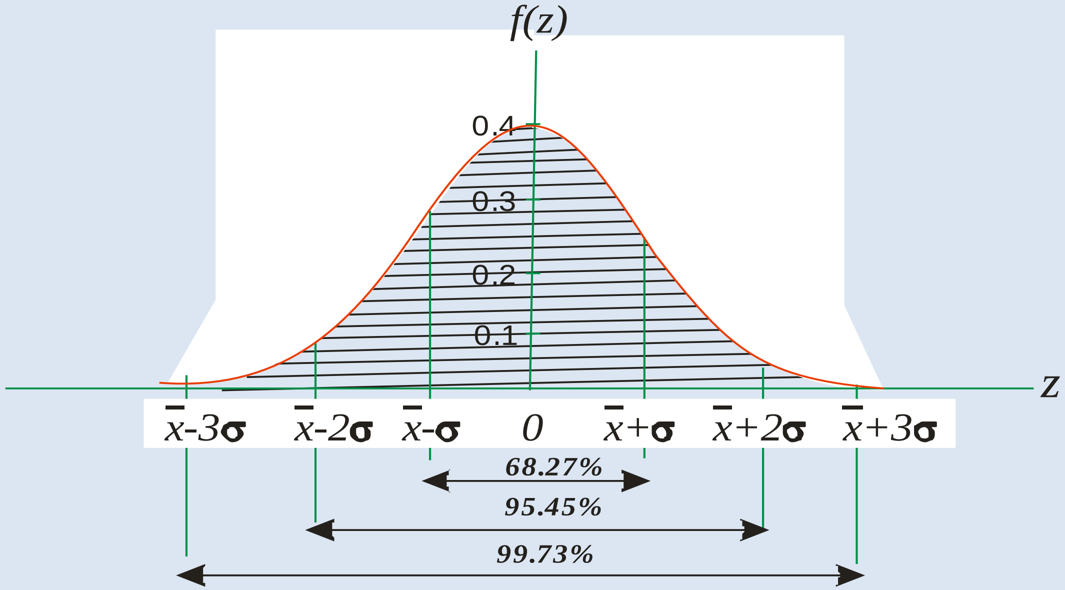
The value z of the standardized variable Z *is sometimes called* the *standard score*.

A graph of the standard normal density function, sometimes called the *standard normal curve*, is shown in Figure below. The areas within 1, 2, and 3 standard deviations of the mean (i.e., between z = -1 and +1, z = -2 and +2, z = -3 and +3) as equal, respectively, to 68.27%, 95.45%, and 99.73% of the total area, which is one. This means that

$P(-1 \le Z \le 1) = 68.27\%$

$P(-2 \le Z \le 2) = 95.45\%$

$P(-3 \le Z \le 3) = 99.73\%$



The table can be used to find areas the curve bounded by the ordinates at z = 0 and any positive value of z, *and also, symmetry of the curve about z = 0 can be used to find areas between any two ordinates.*

Revisiting correlation analysis

BAZE UNIVERSITY ABUJA

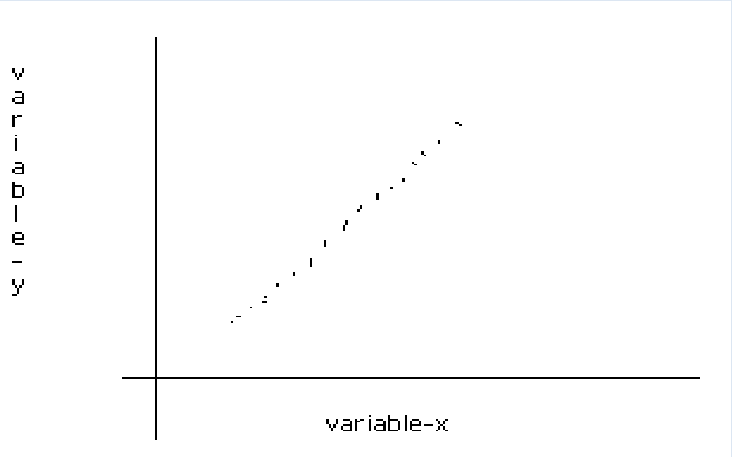
?

Financial data & normal distribution

- ***<u>Note</u>:*** The assumption of homoskedastic and normally distributed data is generally an implausible assumption when using financial data
- Forecasting is a common objective of many econometric / time-series models.
- The objective of a forecast is to minimize a loss function.

Revisiting Correlation Analysis

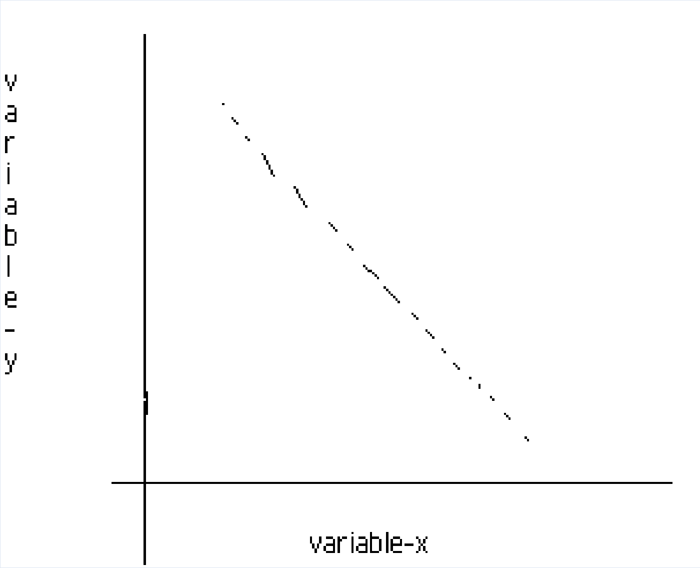
- Correlation is a tool by which we measure the degree of linear relationship between any two quantities such that when this relationship is known, we can then use it to make a forecast about the future, e.g. the extent of the relationship between alcohol consumption and the rates of accidents on our roads.
- Course participants to cite other examples.
- We say that there is correlation between variable x and y if an increase in x leads to a proportional increase or decrease in variable y or vice versa.
- We also say that Perfect Correlation exists between variable x and y if y increases in a correctly definite ratio to an increase in x, the reverse of this gives the perfect negative correlation. This is as represented in the figure below:



Perfect Positive Correlation, r = 1

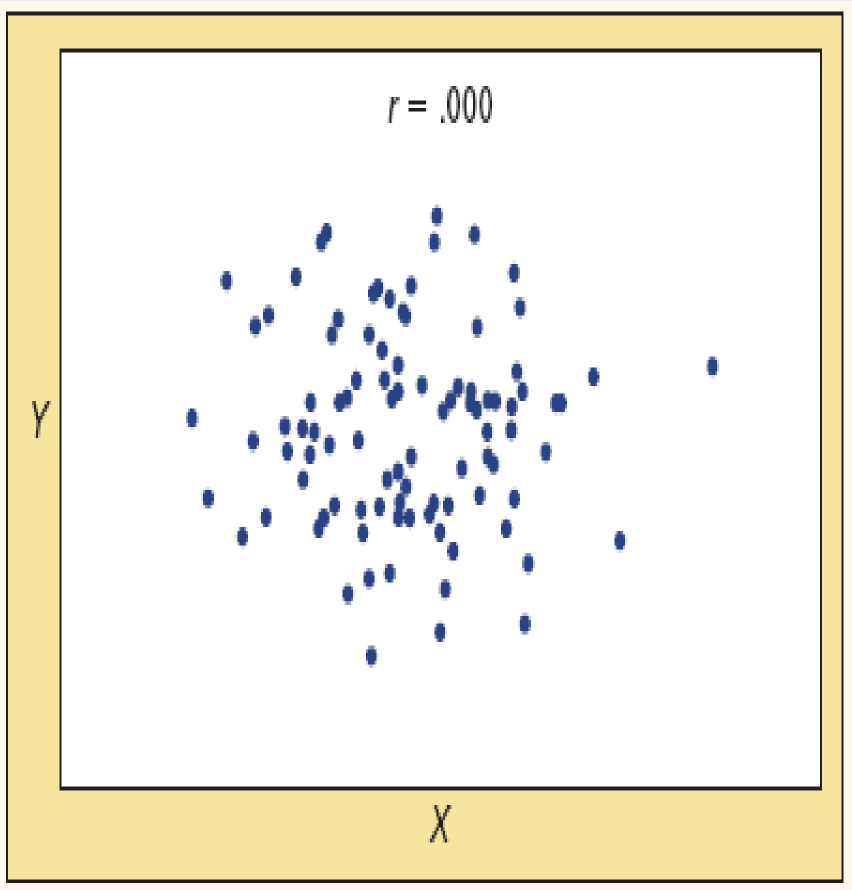
- A practical example two variables that are positively correlated is the price and quantity demanded of a commodity, because when the quantity demanded of a commodity increases, it leads to the proportional increase in its price.

- On the other hand, when two variables x and y tend to change in opposite direction, we say that there is a perfect negative correlation. Here, an increase in value of variable x leads to a proportional decrease in the other variable y. This relationship is as shown in the figure below:



Perfect negative correlation, r = -1

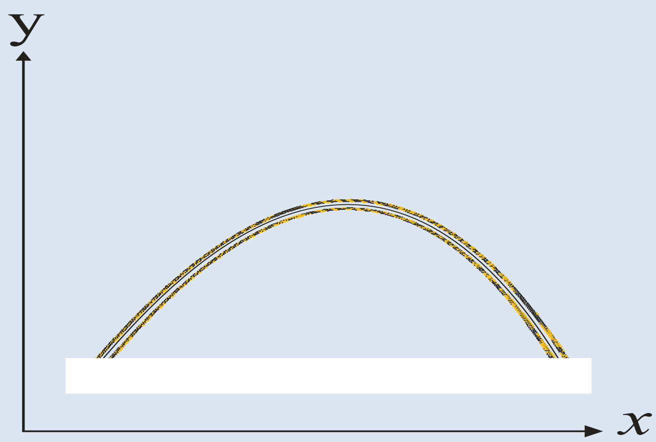
It is said that there is no correlation between any two variables when an increase or decrease in value of one does not have an effect on the other. In such case, the two variables are said to have zero correlation. This is as represented in the figure below:



No correlation, r = 0

- The other form of correlation is the situation where two positively correlated but soon afterwards, the trend change a situation of zero correlation and later becomes that of negat This type of relationship is referred to as curvilinear correla represented in the figure below:

BAZE UNIVERSITY ABUJA



Curvilinear Correlation

- ***REMINDER:*** The degree of correlation that exists between two variables is referred to as Correlation Coefficient, usually denoted by *r* which is a measure of the relationship between variable x and y
- r takes value between -1 and +1.

To compute the Correlation Coefficient, we use any of the following methods:

1. Karl Pearson's Method.
2. Direct method, otherwise called Product – Moment Correlation coefficient method.
3. Spearman's Rank Correlation coefficient method.

Karl Pearson's Method.

To estimate the degree of correlation that exist between two variable x and y we use the formula:

$$r = \frac{\sum\left(x-\bar{x}\right)\left(y-\bar{y}\right)}{\sqrt{\sum\left(x-\bar{x}\right)^2 \sum\left(y-\bar{y}\right)^2}}$$

Where $\bar{x}$ denotes the mean of variable x

and $\bar{y}$ denotes the mean of the second variable y.

Interpretation of correlation coefficient: Summary

BAZE UNIVERSITY ABUJA

S/No	*r*	*Interpretation*
1	= 1	Perfect positive relationship
2	= -1	Perfect negative relationship
3	= 0	No relationship
4	= 0.1 to 0.4	Positive and weak relationship
5	= 0.5 to 0.9	Positive and strong relationship
6	= -0.4 to -0.1	Negative and weak relationship
7	= -0.9 to 0.5	Negative and strong relationship

Revisiting regression analytics…

?

Regression analysis in brief

- The use of equation to establish the relationship between any two variables say X and Y such that if the value of X is known, the corresponding value of Y can be predicted is called **Regression.** Given that two variables x and y are related such that we have function $Y = f(X)$, meaning that the equation connecting X and Y can be written in the form that the functional relationship between X and Y will enable one to predict with exact precision what the corresponding value of y will be.
- **Regression Analysis** is a tool of forecasting through which the analyst can determine the extent to which changes in one variable cause changes in the other and the change itself is estimated.

- **Scatter diagram**
- The values of observations for variable X and that of Y can be plotted on a graph to form what is known as **scatter diagram.**

BAZE UNIVERSITY ABUJA

Interpretation of goodness of goodness of fit: Summary

BAZE UNIVERSITY ABUJA

S/No	***R^2***	***Interpretation***
1	= 1	Perfect fit
2	= 0	No fit
3	= 0.1 to 0.4	Poor fit
4	= 0.5 to 0.9	Good fit

To sum up – common errors in data analytics: Some areas to watch out for

BAZE UNIVERSITY ABUJA

1. <u>Fixation on numbers</u> at the expense of intelligent decision-making - deficient knowledge of what data analysis or statistics is all about – discussion of economic importance or <u>implications</u>
2. Wrong statistical tools/methods – wrong judgment
3. Foggy definition of variables – incomprehensible questionnaires
4. Un-standardized disparity in variable sizes.
5. Assumption of normal distribution viz-a-viz non-normal reality.
6. Using un-scrutinized data, especially secondary data (what Prof Toyin Falola, 2015 called “propagandistic data”, in *The Punch*, 11/6/2015, p. 26)

7. Deficient knowledge of data and variable typologies and how each system should be handled differently in data analytics.

8. Not distinguishing ‘data’ from ‘information’

9. Bias and sampling errors.

10. Fixation on computer outputs - wrong interpretation of analytical results

11. Absence or inadequate training of enumerators - absence of pilot tests.

Revisiting…
Hypothesis testing

BAZE
UNIVERSITY
ABUJA

?

Recall: interpretation of correlation coefficient:

BAZE UNIVERSITY ABUJA

S/No	*r*	*Interpretation*
1	= 1	Perfect positive relationship
2	= -1	Perfect negative relationship
3	= 0	No relationship
4	= 0.1 to 0.4	Positive and weak relationship
5	**= 0.5 to 0.9**	**Positive and strong relationship**
6	= -0.4 to -0.1	Negative and weak relationship
7	= -0.9 to 0.5	Negative and strong relationship

Recall: Interpretation of 'goodness of fit':

BAZE UNIVERSITY ABUJA

S/No	R^2	***Interpretation***
1	= 1	Perfect fit
2	= 0	No fit
3	= 0.1 to 0.4	Poor fit
4	***= 0.5 to 0.9***	***Good fit***

Revisiting the R-squared metric…

BAZE UNIVERSITY ABUJA

- R-Squared is seen as a non-decreasing measure meaning that even as more explanatory variables enter the model, the R-squared value will keep increasing thereby failing to account for the loss of degree of freedom arising from additional independent variables in the model.
- This deficiency is addressed by ***adjusted R-squared*** which gives a truer picture of the goodness of fit of a statistical model.
- Adjusted ***$R^2 > 0.5$*** is thought to imply that the regression model has a good fit and it is therefore okay for forecasting.

Revisiting…
What "<u>statistical significance</u>" really means

BAZE UNIVERSITY ABUJA

- When a statistic is ‘significant’, what this really means in statistics is that you are very sure that the statistic is **<u>reliable</u>**, <u>not</u> that the finding is important or that it has any decision-making utility to your client or audience (unless you can demonstrate this separately in your report).
- Significance is a statistical term that indicates how sure or confident you are that a difference or relationship exists, and how strong or weak it is.
- Significant differences can be large or small depending on your sample size.
- In effect, what we are talking about is not the ordinary sense of significance but ‘statistical significance’.

What 'hypothesis' is all about…

BAZE UNIVERSITY ABUJA

- Recall: ***Hypothesis*** is simply a theory needing investigation: a tentative explanation for a phenomenon, used as a basis for further investigation
- Hypothesis is therefore an assumption or a statement that is assumed to be true for the sake of argument.
- As analysts, we usually want to start our investigations with basic assumptions or *intelligent* guess concerning the behaviour of variables or issue under study.

That market 'assumption' – test it before launching that product!

BAZE
UNIVERSITY
ABUJA

- ***Hypothesis testing*** – testing the validity of your results / claims, scientific verification or validation to confirm whether the assumptions made about the subject variables are true or false.
- Types of hypothesis testing - what do you understand by
  - One-tailed test?
  - Two-tailed test?

On whether to use a <u>one-tailed</u> or <u>two-tailed</u> test of significance?

?

Whether to use a one-tailed or two-tailed test of significance?

BAZE UNIVERSITY ABUJA

- The answer largely depends on your hypothesis.
- To remove biasness, it is generally safest to use two-tailed tests [https://www.statpac.com/surveys/statistical-significance.htm – 29/11/2015]

Hypothesis Testing (HT): Revisiting its essence…

BAZE UNIVERSITY ABUJA

- HT (not High Tension!!) is perhaps an area in which statistics finds its greatest applications – testing validity of claims.
- Testing is needed in practically every field of human activity, particularly business and management research because new ideas and products should be properly evaluated before they are launched into the market place.

Hypothesis testing: continuation…

BAZE UNIVERSITY ABUJA

- Hypothesis tentatively explains an observation that can be tested (i.e. proved or disproved) by further investigation.
- Figuring out the solution to the problem, i.e. "hypothesizing", before you start will help build a roadmap for approaching the problem
- You can express hypothesis as possible root causes of the problem
- Breaking down the problem into key drivers (root causes) or aspects can help formulate hypothesis.

Formulating your hypothesis…

BAZE UNIVERSITY ABUJA

❑ Null Hypothesis: H_o

- Ho is set up by the investigator with the intention of being rejected based on the available statistical evidence.

❑ Alternative Hypothesis: H_1

- H_1 is the hypothesis accepted by the investigator after the Ho has been rejected.

Type I and Type II decision errors

BAZE UNIVERSITY ABUJA

❑ *Type I error – alpha (α) error:*

- True H_o is rejected
- False H_1 is accepted

❑ *Type II error – beta (β) error:*

- True H_1 is rejected
- False H_o is accepted
- Why is Type II error usually emphasized in statistical literature?

Formal testing statistics…

BAZE UNIVERSITY ABUJA

- ❑ <u>*t*-test</u> statistics (the student *t*-test) – small sample test, i.e. $n < 30$
- ❑ <u>*Z*-test</u> statistics (normal test) – large sample test i.e. $n \geq 30$
- ❑ <u>*F*-test</u> statistics (joint test)
- ❑ <u>Chi-square</u> distribution X^2

Formal 'confidence levels'…

BAZE UNIVERSITY ABUJA

- Confidence level – the extent of confidence or certainty that the investigator has; conventionally (2-tailed):
- $Z_T = Z\alpha$ = 1% = 2.58
- **$Z_T = Z\alpha$ = 5% = 1.96 (This is the widely used range in social & management sciences)**
- $Z_T = Z\alpha$ = 10% = 1.65

********Decision Rule:***

- Calculated test statistic > tabular (critical) value: Accept H_1 and reject H_o

i.e. the parameter testes is statistically significant

- Calculated test statistic < tabular (critical) value: Accept H_o and reject H_1

i.e. the parameter testes is statistically insignificant

*****Statistical testing:

On simpler (less confusing!) approaches…

BAZE UNIVERSITY ABUJA

- The *p*-value is the probability that the test statistic (*z**, *t**, *x**, *F**) will be exceed, and thus *p* is called <u>*the observed level of significance*</u>, in contrast to the α-value which is a priori-level of significance.
- The default value of $\alpha = 0.05$, and the relationship between P and α is as stated below (Kothari & Garg, 2014):

i. **If $p \geq \alpha$, do <u>not</u> reject Ho**

ii. **If $p < \alpha$, reject Ho**

Statistical significance test:

On simpler (less confusing!) approaches…

BAZE UNIVERSITY ABUJA

- In essence, using a table is not necessary <u>when</u> you have the exact probability for a statistic.
- Your econometric software can calculate exact probabilities for most test statistics. If you have an exact probability output from computer software, you simply compare it to your critical alpha level.
- If the exact probability is less than the critical alpha level, your finding is significant (i.e. Ho rejected)
- if the exact probability is greater than your critical alpha level, your finding is <u>not</u> significant (i.e. Ho accepted)

Formal statistical 'significance' test:

To sum up…

BAZE UNIVERSITY ABUJA

S/No	Observation	Interpretation rule
1	If calculated t-statistic < 1.96 (i.e. 5% significance level)	Accept Ho
2	If calculated $p > 0.05$ (i.e. 5% significance level)	Accept Ho

Type I and Type II decision errors…

BAZE UNIVERSITY ABUJA

- Pick hypothesis before looking at the data to avoid bias – *apriori* expectation
- Don't confuse statistical importance with substantive importance.
- The *p*-value is not necessarily the probability that the null hypothesis (*Ho*) is true; the *p*-value already assumes that *Ho* is true.
- Rather, it is the probability of rejecting Ho incorrectly on the basis of your results that is displayed in sample assessment.

- You have learnt about how to interpret the adequacy of statistical models, such as beta coefficients, R-squared, adjusted R-squared, finding the critical value of *F* statistic and verifying the *F* calculated value, but ensure that you able to summarize your results in everyday language – less jargons, please!

Hypothesis testing
The standard procedure revisited…

BAZE UNIVERSITY ABUJA

I
Formulate Hypotheses

II
Select confidence level

III
Select estimator

IV
Select tail type

Hypothesis testing
The standard procedure revisited…

V

Calculate test statistic

VI

Compare t-value to the rejection region

VII

Make your conclusions

Statistical testing:
…made easier by technology, but…

BAZE UNIVERSITY ABUJA

- ❑ The manual procedure for hypothesis testing can be tedious.
- ❑ Happily, there are statistical tools available these days to solve the problems with relative ease, once they have been carefully / properly defined.
- ❑ All you need to do is enter the given data and then issue the appropriate commands.
- ❑ Software supporting hypothesis tests: SPSS, MINITAB, EViews, MS Excel, and so on.
- ❑ When the results of manual solutions are compared to the computer solutions, we see that they are identical.
- ❑ But…

some provisos…

BAZE UNIVERSITY ABUJA

- ✓ Understand the methodology (which we described earlier).
- ✓ Be familiar with the output generated by the computer (you will see some examples of these in the practical session).
- ✓ Be knowledgeable enough to interpret the computer results meaningfully (<u>again, the recurring imperative of proper understanding of the related theory, the business and the question at hand</u>)

Introduction to time series analysis

A Conceptual Overview

Spurious correlation…

BAZE UNIVERSITY ABUJA

“First impressions can often lead to wrong conclusions.”

-*Our Daily Bread*, 2015, p. 346.

Spurious Correlation…

BAZE UNIVERSITY ABUJA

- Correlation does not mean causation
- Consequently regression may be also spurious and interpretation may not be valid.

…Non-stationary time series

Basic classifications / sources:

1. Seasonality – pervasive in economic time series
2. Deterministic trends (time trends)
3. Unit Roots (Stochastic trends)
4. Structural breaks

- ***Seasonality*** – data exhibiting a non-constant deterministic pattern with an annual frequency.
- ***Hebdomadality*** – data exhibiting day-of-the-week deterministic effects.
- ***Diurnality*** – Data that exhibit intra-day deterministic effects

Spurious regression…

BAZE UNIVERSITY ABUJA

- According to Kumar (2015), when nonstationary time series are used in a regression model, the results may spuriously indicate a significant relationship when there is none
- In these cases the least squares estimator and least squares predictor do not have their usual properties, and *t*-statistics are not reliable
- Since many macroeconomic and financial time series are nonstationary, it is crucial to exercise care when estimating regressions with dynamic variables.

Still on spurious regression…

BAZE UNIVERSITY ABUJA

- Usually the regression only tells us there is <u>some</u> 'relationship' between *x* and *y*, and does not tell the nature of the relationship, such as <u>*whether x causes y or y causes x*</u>.
- Granger test can be used in investigating whether or not *Y* causes *X*. (Kumar, 2015)

Understanding time-based data…

- In time-based data, the change in a variable is an important concept.
- The change in a variable y_t, also known as its first difference, is given by $\Delta y_t = y_t - y_{t-1}$.
- Δy_t is the change in the value of the variable y from period t - 1 to period t

Understanding time series analysis…

BAZE UNIVERSITY ABUJA

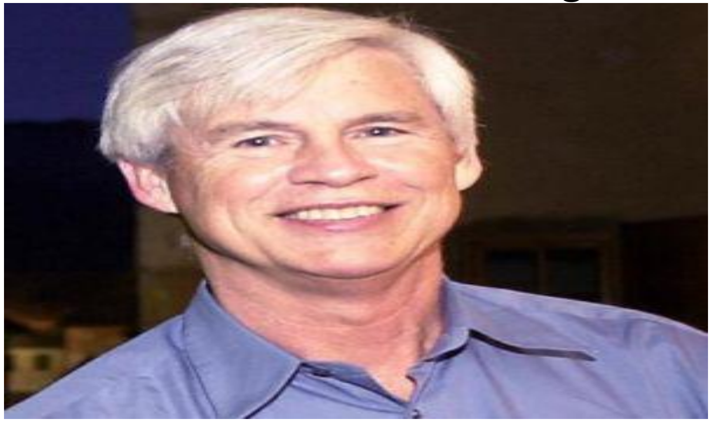
- Time-based data – showing the dynamic movement of a phenomenon over a period of time, usually at equal intervals.
- Used in any domain of applied sciences and engineering that involve temporal measurements.
- Sequence of data points – successive movements over time, e.g. ocean tides, closing stock prices on Indian stock markets, etc.

Robert F. Engle & Sir Clive William John Granger

BAZE UNIVERSITY ABUJA

- Sir Clive William John **Granger** was a British economist, who taught in Britain at the University of Nottingham and in the United States at the University of California, San Diego. In 2003, Granger was awarded the Nobel Prize in Economic Sciences in recognition of what he and his co-winner, Robert F. **Engle** (picture below) had made contributions to the analysis of time series data that had changed fundamentally the way in which economists analyse financial and macroeconomic data, enabling economists to make more accurate stock-market forecasts.
- Engle co-founded the ***Society for Financial Econometrics*** (SoFiE), a network of experts committed to the field of financial econometrics. In 2009, he founded the ***Volatility Institute*** at NYU. This organization promotes research on the issue of risk in financial markets. Engle also directs the NYU V-Lab, which provides forecasting and analysis of market trends using both classic models and newer tools.
- [http://www.biography.com/people/robert-f-engle-13607503#major-contributions - 28/11/2015]

Sir Clive William John Granger



Stationarity & its essence

BAZE UNIVERSITY ABUJA

- ***Stationarity*** - a concept / tool aimed at transforming raw data to become stationary (so that joint probability distribution parameters such as mean and variance do not change when shifted in time).
- ***Stationary time series*** is one whose statistical properties such as mean, variance, autocorrelation, and so on are all constant over time.
- *****The world of science is cautious about trying to extrapolate *regression* models fitted to non-stationary data.

<u>*Stationarity & its essence*</u>

BAZE UNIVERSITY ABUJA

- ***Stationarity*** acknowledges stochastic processes
- A stochastic process is an arbitrary sequence of random data
- Random walk: $y_t = y_{t-1} + \epsilon_t$
- Hence, realizing randomness of time series, stationarity is a measure of regularity exploited to allow us to make estimation of unknown parameters and characterize the dependence between observations across time.
- <u>If dataset is allowed to change in an unpredictable manner, constructing a meaningful model would be difficult or impossible.</u>

Stationarity - continuation…

BAZE UNIVERSITY ABUJA

- Formally, a time series y_t is stationary if its mean and variance are constant over time, and if the covariance between two values from the series depends only on the length of time separating the two values, and not on the actual times at which the variables are observed (Kumar, 2015).
- That is, the time series y_t is stationary if for all values, and every time period, it is true that:

$E(y_t) = \mu$ (constant mean)

$\mathrm{var}(y_t) = \sigma^2$ (constant variance)

$\mathrm{cov}(y_t, y_{t+s}) = \mathrm{cov}(y_t, y_{t-s}) = \gamma_s$ (covariance depends on s, not t)

Explaining stationarity (Kumar, 2015) continuation

BAZE UNIVERSITY ABUJA

- The autoregressive model of order one, the AR(1) model, is a useful univariate time series model for explaining the difference between stationary and non-stationary series: $y_t = \rho y_{t-1} + v_t, \quad |\rho| < 1$
  - The errors v_t are independent, with zero mean and constant variance, and may be normally distributed
  - The errors are sometimes known as "shocks" or "innovations"

Explaining stationarity (Kumar, 2015) continuation

BAZE UNIVERSITY ABUJA

- <u>The main reason why it is important to know whether a time series is stationary or non-stationary before one embarks on a regression analysis is that there is a danger of obtaining apparently significant regression results from unrelated data when non-stationary series are used in regression analysis.</u>
- Such regressions are said to be **spurious**
- Assume two independent random walks:

$$
\begin{aligned}
rw_1 &: \quad y_t = y_{t-1} + v_{1t} \\
rw_2 &: \quad x_t = x_{t-1} + v_{2t}
\end{aligned}
$$

  - These series were generated independently and, in truth, have no relation to one another, yet this may not be apparent from its graph (see next slide)

Understanding the rationale behind stationarity tests…

BAZE UNIVERSITY ABUJA

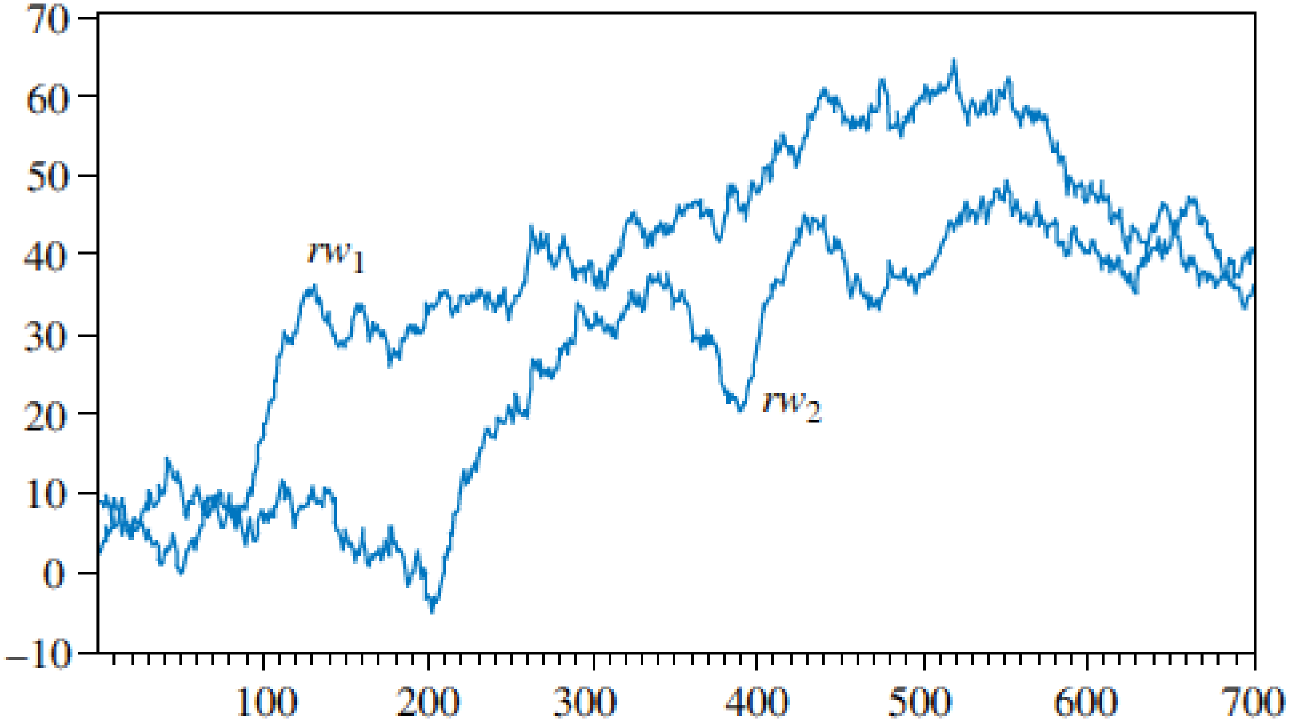
- ❑ We need to stationarize a time series to be able to obtain meaningful sample statistics such as means, variances, and correlations with other variables.
- ❑ ***Such statistics are useful as descriptors of future behaviour only if the series is stationary***.

- ***Other motivations:***

1. Most statistical forecasting methods are based on the assumption that the time series can be rendered approximately stationary (i.e., "stationarized") through the use of mathematical transformations.
2. A stationarized series is relatively easy to predict: you simply predict that its statistical properties will be the same in the future as they have been in the past. Your computer software normally takes care of the requisite computations and transformation details.
3. Stationarizing a time series through differencing (if required) is an important part of the process of fitting an ARIMA model. (Hatemi, 2004).

Still on stationarity (Kumar, 2015) continuation

BAZE UNIVERSITY ABUJA



(*a*) Time series

Time series analysis: Understanding other key terms…

BAZE UNIVERSITY ABUJA

- ***Cointegration*** (long-run relationship between two moving variables) is a statistical property of a collection (X1,X2,...,Xk) of time series variables. First, all of the series must be integrated of order 1. Thereafter, if a linear combination of this collection is integrated of statistical order zero, then the collection is said to be co-integrated.
- ***Autocorrelation*** - a mathematical tool for finding repeating patterns, such as the presence of a periodic signal obscured by noise – serial dependence

- ***Lags:*** a period of time between one event and another.
- ***Random walk*** – each value is completely a random step away from the previous value (not auto-correlated)

Understanding ‘differencing’

BAZE UNIVERSITY ABUJA

- ***Differencing*** – this is a viable method of transforming a nonstationary series to become stationary
- ***First difference*** of a time series is the series of changes from one period to the next. If Y_t denotes the value of the time series Y at period t, then the first difference of Y at period t is equal to $Y_t - Y_{t-1}$.
- First difference is useful filter to separate a “trend” from “cyclic” component in a series.
- Don’t over-difference – applying the difference operator to a stationary series.

Understanding Lags… (not Lagos!!)
& Autocorrelations…

BAZE UNIVERSITY ABUJA

- **Distributed lag model** is a model for time series data in which a regression equation is used to predict current values of a dependent variable based on both the current values of an explanatory variable and the **lagged** (past period) values of this explanatory variable.
- ***Autocorrelations*** are to autocovariances as correlations are to covariances.
- The ***autocorrelation function (ACF)*** relates the ***lag length*** and the parameters of the model to the autocorrelation.
- The ACF is a function of the population parameters that defines the relationship between the autocorrelations of a process and lag length.

Still on Lags…

BAZE UNIVERSITY ABUJA

- The standard practice when working with non-stationary / seasonal data is to conduct model selection over ***two sets of lags*** by choosing a maximum lag to capture the seasonal dynamics and by selecting a maximum lag to capture non-seasonal ones

On time series models dealing with conditional variances

BAZE UNIVERSITY ABUJA

- Dealing with ***non-normal*** /non-linear distributions and ultra-high frequency data. - the availability of more and better data and the availability of low-cost high-performance computers allowed the development of a vast family of ***ARCH/GARCH models*** originally developed by Robert F. Engle (1982) [in his "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation". *Econometrical* 50, 4: 987–1007]
- While forecasting of expected returns perhaps still remains a rather elusive task, analyzing and predicting the level of uncertainty and the strength of co-movements between asset returns has become a fundamental pillar of financial econometrics.

- ***Rationale:***
  - Precision analysis: More precise analysis and forecasting.
  - Towards finding steady investment strategies.
  - Determining the usefulness and reliability of trading strategies.
  - Increased usage where volatility of returns is a key issue, e.g. with electronic trading comes massive amount of data – measurement of intraday risk and discovery of trading profit opportunities
  - There are phenomena that exist at some time horizon and disappear at other time horizon.

Nonlinear models for financial time series analysis….

BAZE UNIVERSITY ABUJA

- ❑ Autoregressive models **(AR)**
- ❑ Integrated models **(I)**
- ❑ The moving average models **(MA)**
- **ARMA:** Autoregressive moving average
- **ARIMA:** Autoregressive integrated moving average. In time series analysis, ARIMA model is a generalization of ARMA model. These models are fitted to time series data set either to better understand the data or to forecast – i.e. to predict future points in the series. They are applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the "integrated" part of the model) can be introduced to reduce the non-stationarity.
- **ARCH:** Autoregressive conditional heteroskedasticity
- **GARCH:** Generalized autoregressive conditional heteroskedasticity
- **TARCH –** Threshold autoregressive conditional heteroskedasticity

- ❑Markov Switching Autoregression (MSAR)
- ❑Threshold Autoregression (TAR)
- ❑Self-Exciting Threshold Autoregression (SETAR).

On time-series models dealing with conditional variances

BAZE UNIVERSITY ABUJA

S/No	Model	Rationale
1	ARCH	Models volatility; deals with time-varying variances (heteroskedasticity) that depend on lagged effects (autocorrelation) commonly observed in many financial market variables.
2	GARCH	Captures long-lagged effects of fewer parameters. This fits many data series well.
3	TARCH	Treats positive and negative news asymmetrically.
4	ANN	Artificial Neural Networks – dealing with large, complex, non-linear or interconnected data sets that are hard to solve using conventional approaches

Models for time series data (continuation)

BAZE UNIVERSITY ABUJA

- ***<u>Error Correction Model (ECM</u>***) is a theoretically-based time series models that directly estimates the speed at which a dependent variable (Y) returns to equilibrium after a change in an independent variable (X). ECMs are useful for estimating both short-term and long-term effects of one time series on another when dealing with stationary data and co-integrated data.
- The **<u>Vector Autoregression (VAR)</u>** is used to capture the <u>linear interdependencies</u> among multiple time series VAR models generalize the univariate autoregressive (AR model) by allowing for more than one evolving variable.
- A **VAR model** describes the evolution of a set of *k* variables (called endogenous *variables*) over the same sample period (t = 1... T) as a <u>linear function</u> of only their past values. The variables are collected in a $k \times 1$ vector y_t, which has as the i th element, $y_{i,t}$, the observation at time "t" of the i th variable. For example, if the i th variable is GDP, then $y_{i,t}$ is the value of GDP at time t.

Model selection:

The Box-Jenkins Methodology

BAZE UNIVERSITY ABUJA

- The Box-Jenkins methodology is one of the widely used approaches in financial econometrics
- Has two aspects:
  - ***Identification*** – visual inspection of the series – the autocorrelations and partial autocorrelation
  - ***Estimation*** – candidate models are identified by relating the sample autocorrelations and partial autocorrelations to the autocorrelation function and partial autocorrelation function of the ARMA models

- ***The BJ procedure relies on two principles:***

i. ***Parsimony*** – the specification with the fewest parameter capable of capturing the dynamics of a time series is preferred to other representations equally capable of capturing the same dynamics.

ii. ***Invertibility*** – a technical requirement stemming from the use of the autocorrelogram and partial autocorrelogram to choose a model – achieving a unique identification of the moving average component of a model.

Granger causality test

BAZE UNIVERSITY ABUJA

- ❑In time series analysis, you would like to know whether changes in a variable will have an impact on changes other variables.
- ❑ Granger causality test is used to address such phenomena more accurately.

Granger causality, yes but…

BAZE UNIVERSITY ABUJA

- ***Granger causality*** does not equal to what is usually meant by causality.
- Even if A does not 'cause' B, (in the ordinary sense of the word 'cause'), it may still help to predict B, and thus Granger-causes B if changes in A precedes that of B for various reasons.

Granger Causality continuation…

BAZE UNIVERSITY ABUJA

- In principle, if *X* causes *Y*, then, changes of *X* happened <u>first</u> then followed by changes of *Y*.
- If *X* causes *Y*, there are <u>two conditions</u> to be satisfied:

1. *X* can help in predicting *Y*. Regression of *X* on *Y* has a big R^2
2. *Y* can not help in predicting *X*.

Granger Causality continuation…

BAZE UNIVERSITY ABUJA

- In the context of two variables, *x* and *y*, *y* is said to Granger-cause *x* if current or lagged values of *y* helps to predict future values of *x*.
- On the other hand, *y* fails to Granger-cause *x* if for all s > 0, the mean squared error of a forecast of x_{t+s} based on $(x_t, x_{t-1}, \ldots)$ is the same as that is based on $(y_t, y_{t-1}, \ldots)$ and $(x_t, x_{t-1}, \ldots)$.

Understanding causality test: Two broad possibilities

BAZE UNIVERSITY ABUJA

- ❑ Unidirectional causality
- ❑ Bidirectional causality
- Uni-directionality: That all millionaires were persistent, hardworking people does not make hard workers become millionaires (does it?); plenty of failed entrepreneurs were persistent, hard working people.
- Similarly, risk-taking is necessary for large success, but it is also necessary for failure. [Nassim Nicholas Taleb, 2004].

Causality test
<u>NOTE</u>: 4 possible specific outcomes…

1. *X* Granger causes *Y* but *Y* does not Granger cause *X*
2. *Y* Granger causes *X* but *X* does not Granger cause *Y*
3. *X* Granger causes *Y* and *Y* Granger causes *X* (i.e., there is a <u>feedback system</u>)
4. *X* does not Granger cause *Y* <u>and</u> *Y* does not Granger cause *X*

Granger causality testing procedure

BAZE
UNIVERSITY
ABUJA

?

Granger causality testing procedure:
Three conditions (Kumar, 2015)

BAZE UNIVERSITY ABUJA

1. Establish correlation first.
2. There must be issue of timing – the independent variable must have changed for the dependent variable to react
3. Third or other factors must be isolated.

Granger causality testing process…

- Stationary and non-stationary variables
- Spurious regressions
- Unit root tests - for non-stationarity
- Cointegration
- When there is no cointegration – what do you do?
- Granger’s causality test

Order of integration of the variables

BAZE UNIVERSITY ABUJA

- Note that all variables have to be of the same order of integration; the following are possible cases (Hatemi, 2004):

1. *All the variables are I(0) (stationary):* one is in the standard case, i.e. a VAR in level.
2. All the variables are I(*d*) (non-stationary) with *d* > 0
3. **The variables are co-integrated: the error correction term has to be included in the VAR. The model becomes a Vector Error Correction Model (VECM).**
4. The variables are not *co-integrated* : the variables have first to be differenced *d* times and one has a VAR in difference.

Granger causality testing process continuation…

BAZE UNIVERSITY ABUJA

STEPS	*PURPOSE*
Graphs	Visualization to 'see' if there is any possibility of correlation to begin with.
Unit Root Test (Augmented Dickey Fuller method is commonly used)	Testing for non-stationarity to ensure the validity of empirical results. Note: We can test for non-stationarity by testing the null hypothesis that ρ = 1 against the alternative that |ρ| < 1 (Kumar, 2015)
Co-integration (Johansen System Cointegration test is commonly used)	Testing for short-run relationship between two moving variables. The testing statistic is $\tau = {}^{\hat{\varphi}}/_{se\ (\hat{\varphi})}$
Vector Error Correction Model (VECM)	Fitting an error correction model if co-integration is established - to check whether error correction mechanism takes place if some disturbance comes in the equilibrium relationship, i.e. to measure the speed of convergence to the long-run steady state of equilibrium
Granger Causality	Establishes presence of causality and its direction – i.e. to examine if the correlation coefficients have causal relationship

What Unit Root Test is all about

- A primer

What do we mean by 'unit root' and 'unit root test' (URT)?

BAZE UNIVERSITY ABUJA

- A **URT** is a <u>statistical tes</u>t for the idea or proposition that in a autoregressive statistical model of a time series, the autoregressive parameter is one.
- A <u>unit root is an attribute</u> of a statistical model of a time series whose autoregressive parameter is one.
- *In $y_t = y_{t-1} + ut$, where $-1 \leq p \leq 1$, and ut is a white noise error term, **<u>if p is in fact 1</u>**, then we face what is called **<u>the unit root problem,</u>** i.e. a situation of non-stationarity (a non-stationary stochastic process).*

Checking for Unit Root in time series data…

BAZE UNIVERSITY ABUJA

- Recall that if p = 1, we have a unit root problem, i.e. a non-stationary stochastic process.
- We check for presence of unit root by regressing yt on its lagged value y_{t-1} to establish whether the calculated *p* is equal to 1.
- If $p = 1$, then *yt* (i.e. the time series under analysis) is non-stationary.
- Note: If $p = 0$, then *yt* is stationary (i.e. the first differences of a random walk time series are stationary).

Obtaining the right order of integration:

Unit Root Test (URT) for stationarity

BAZE UNIVERSITY ABUJA

- ***Unit root processes*** are generalizations of the classic random walk – described mathematically earlier
- A process is said to have a unit root if the distribute lag polynomial can be factored so that one of the roots is exactly one.
- Many economic / financial time series have roots close to 1; thus, it is important to maximize the power of a unit root test so that models posses the right order of integration.

Types of Unit Root Test

BAZE UNIVERSITY ABUJA

1. Dickey-Fuller **(DF)**
2. Augmented Dickey-Fuller **(ADF)**
3. Phillip Perron **(PP)** – modification of ADF

- ***Note:*** In time series analysis, we have to specify which model of the three URT models we wish to use.
- We also need to specify the number of lagged dependent variables to be included in the model in order to correct the presence of serial correlation.

Unit Root Test of Stationarity Dickey–Fuller test

BAZE UNIVERSITY ABUJA

- The most popular URT is the Dickey–Fuller test - You want to test whether ρ is equal to one or significantly less than one
- The AR(1) process $y_t = \rho y_{t-1} + v_t$ is stationary when $|\rho| < 1$
- But, when $\rho = 1$, it becomes the non-stationary random walk process

Unit Root Test (URT) of stationarity Dickey–Fuller method…

BAZE UNIVERSITY ABUJA

- The most popular URT is the ***Augmented* Dickey–Fuller** (ADF) test - You want to test whether ρ is equal to one or significantly less than one

- ***NOTE:*** The AR(1) process $y_t = \rho y_{t-1} + v_t$ is stationary when **$|\rho| < 1$**
- But, when **$\rho = 1$**, it becomes the non-stationary random walk process

Unit Root Test of stationarity
Dickey–Fuller test

BAZE UNIVERSITY ABUJA

- The most popular URT is the Dickey–Fuller test - **You want to test whether ρ is equal to one or significantly less than one**
- The AR(1) process $y_t = \rho y_{t-1} + v_t$ is stationary when $|\rho| < 1$
- But, when $\rho = 1$, it becomes the non-stationary random walk process

DF Test: The hypotheses

BAZE UNIVERSITY ABUJA

- Consider the AR(1) model:

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

We can test for non-stationarity by testing the null hypothesis that ρ = 1 against the alternative that |ρ| < 1 (Kumar, 2015).

DF Test: The hypotheses continuation…

BAZE UNIVERSITY ABUJA

- An alternative format is:

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

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

$$= \gamma y_{t-1} + v_t$$

  – The **hypotheses** are:

$$H_0 : \rho = 1 \quad \Leftrightarrow \quad H_0 : \gamma = 0$$

$$H_1 : \rho < 1 \quad \Leftrightarrow \quad H_1 : \gamma < 0$$

The Dickey-Fuller testing procedure

BAZE UNIVERSITY ABUJA

- First plot the time series of the variable and select a suitable Dickey-Fuller test based on a visual inspection of the plot
- If the series appears to be wandering or fluctuating around a sample average of zero, use test equation **(a)**
- If the series appears to be wandering or fluctuating around a sample average which is non-zero, use test equation **(b)**
- If the series appears to be wandering or fluctuating around a linear trend, use test equation **(c)**

DF Test with constant but no trend

BAZE UNIVERSITY ABUJA

- The second Dickey–Fuller test includes a constant term in the test equation:

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

The null and alternative hypotheses are the same as before

- The third Dickey–Fuller test includes a constant and a trend in the test equation:

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

–The null and alternative hypotheses are $H_0: \gamma = 0$ and $H_1: \gamma < 0$

DF Test Critical value

BAZE UNIVERSITY ABUJA

- To test the hypothesis in all three cases, we simply estimate the test equation by least squares and examine the *t*-statistic for the hypothesis that

  $\gamma = 0$

  – Unfortunately this *t*-statistic no longer has the *t*-distribution

  – Instead, we use the statistic often called a τ (*tau*) statistic (Kumar, 2015).

Critical values (Kumar, 2015)

BAZE UNIVERSITY ABUJA

To carry out a one-tail test of significance, if τ_c is the critical value obtained from Table, we reject the null hypothesis of non-stationarity if $\tau \leq \tau_c$

If $\tau > \tau_c$ then we do not reject the null hypothesis that the series is non-stationary

Model	1%	5%	10%
$\Delta y_t = \gamma y_{t-1} + v_t$	−2.56	−1.94	−1.62
$\Delta y_t = \alpha + \gamma y_{t-1} + v_t$	−3.43	−2.86	−2.57
$\Delta y_t = \alpha + \lambda t + \gamma y_{t-1} + v_t$	−3.96	−3.41	−3.13
Standard critical values	−2.33	−1.65	−1.28

Note: These critical values are taken from R. Davidson and J. G. MacKinnon (1993), *Estimation and Inference in Econometrics*, New York: Oxford University Press, p. 708.

Mathematics of unit root test (Kumar, 2015)

- Hence, testing for a ***unit root*** is equivalent to testing ϕ=1 in the following model

$$ADF\ test\ equation: Y_t = \phi Y_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta Y_{t-j} + \theta_0 + a_t$$

$$\Delta Y_t = \underbrace{(\phi - 1)}_{\delta} Y_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta Y_{t-j} + \theta_0 + a_t$$

$$ADF\ test\ equation: \Delta Y_t = \delta Y_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta Y_{t-j} + \theta_0 + a_t$$

$$H_0 : \phi = 1 \qquad H_0 : \delta = 0$$

$$H_1 : |\phi| < 1 \qquad H_1 : \delta < 0$$

Reject H_0 if $t_{\theta=1}$<CV

Reject H_0 if $t_{\delta=0}$<CV

Usage of Granger test: A brief Illustration (Kumar, 2015)

BAZE UNIVERSITY ABUJA

- ***World Oil Price and Growth of US Economy – a study by James Hamilton*** (using 91 – 95 observations)
- **The enquiry:** Does the increase of world oil price influence the growth of US economy or does the growth of US economy affect the world oil price?

- There are two causalities that need to be observed:

(i) H_0: Growth of US Economy does not influence world oil price

(ii) H_0 : World oil price does not influence growth of US Economy

- ***James Hamilton's F Tests Results:***

1. Hypothesis that US economy does not effect world oil price is not rejected. It means that the US economy does not have effect on world oil price.
2. Hypothesis that world oil price does not influence US economy is rejected. It means that the world oil price does influence US economy.

Again: Chicken vs. Egg

Which comes first?

?

Chicken vs. Egg Which comes first?

BAZE UNIVERSITY ABUJA

- Thurman and Fisher (1988) using yearly data of chicken population and egg productions in the USA from 1930 to1983 concludes that:
  1. Egg causes the chicken.
  2. There is no evidence that chicken causes egg.
- Hypothesis that egg has no effect on chicken population is thus rejected; while the other hypothesis that chicken has no effect on egg is not rejected.
- Why?

On chicken vs. egg continuation…

BAZE UNIVERSITY ABUJA

- Thurman and Fisher (1988) using yearly data of chicken population and egg productions in the USA from 1930 to1983 concludes that:

  1. Egg causes the chicken.

  2. There is no evidence that chicken causes egg.

- Hypothesis that egg has no effect on chicken population is thus rejected; while the other hypothesis that chicken has no effect on egg is not rejected.
- Why?

Further application: Efficient Market Hypothesis (EMH)

BAZE UNIVERSITY ABUJA

- The ***Efficient Market Hypothesis (EMH),*** which suggests that *returns of a stock-market are unpredictable from historical price changes*, is satisfied when stock prices are characterized by a random walk (***unit root***) process.
- A finding of unit root implies that stock returns cannot be predicted (Munir *et al,* 2012)

*****Applications: Further perspectives…

BAZE UNIVERSITY ABUJA

- If two variables are co-integrated, in the long-run these two variables will have a common trend. If we have market efficient efficiency, co-integration must exist in the futures and spot market. This means that the futures prices will not be consistently above or below the spot prices (see Hakkio & Mark, 1989, in Market efficiency and co-integration: An application to the sterling and Deutschmark exchange markets, *Journal of International Money and Finance*, 8, pp. 75-88)
- If co-integration exists between two variables, this means that the futures prices can be used to predict spot prices. (see Granger, 1986, in Developments in the study of co-integrated economic variables, *Oxford Bulletin of Economics and Statistics*, 48, pp. 213 - 228)

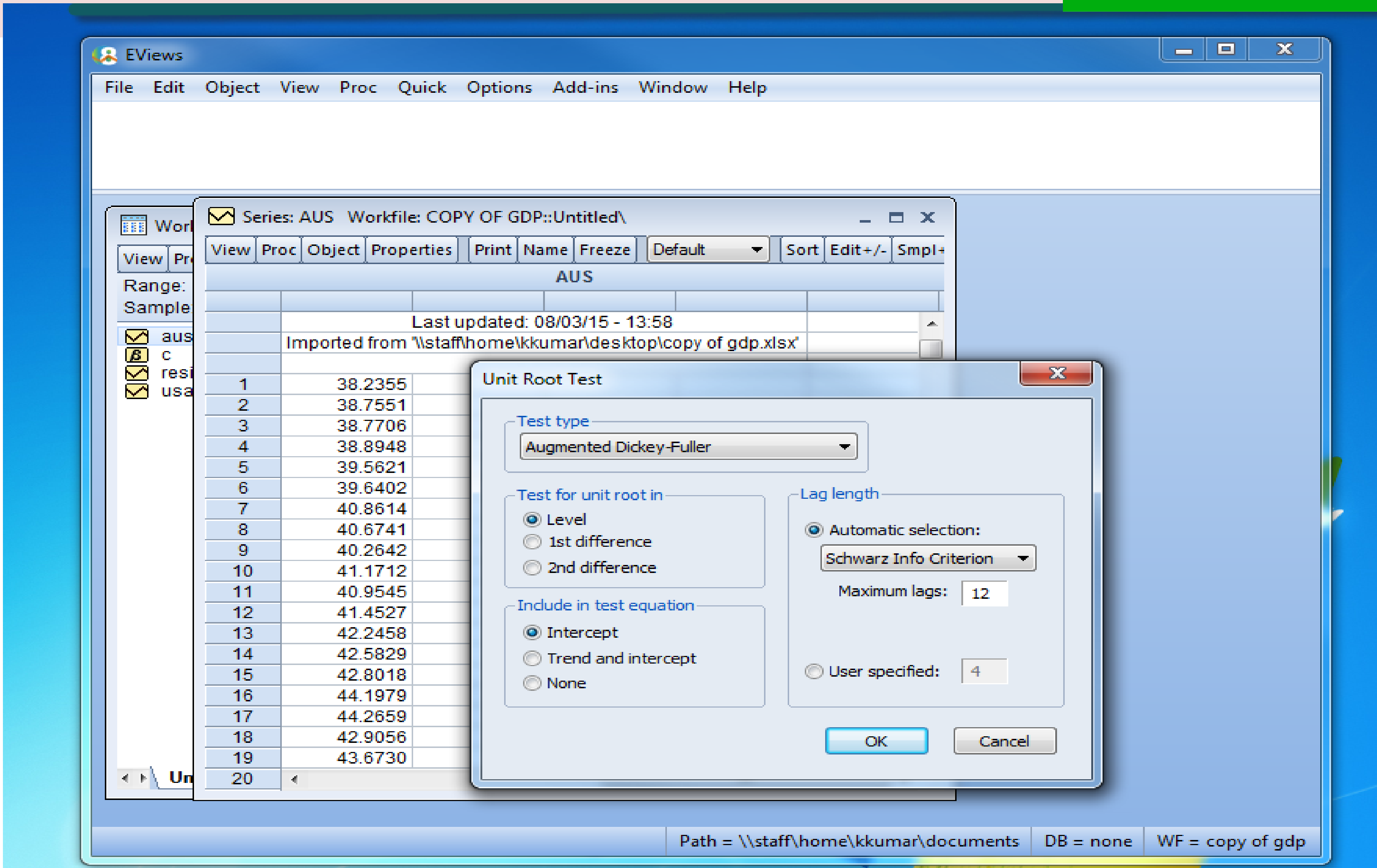
Applications: Further perspectives…

BAZE UNIVERSITY ABUJA

- The next few slides provide further causality illustrations provided by Professor Kumar (2015)based on this enquiry:
- *Does the US economy influence Australia economy or does the Australia economy influence the US economy?*

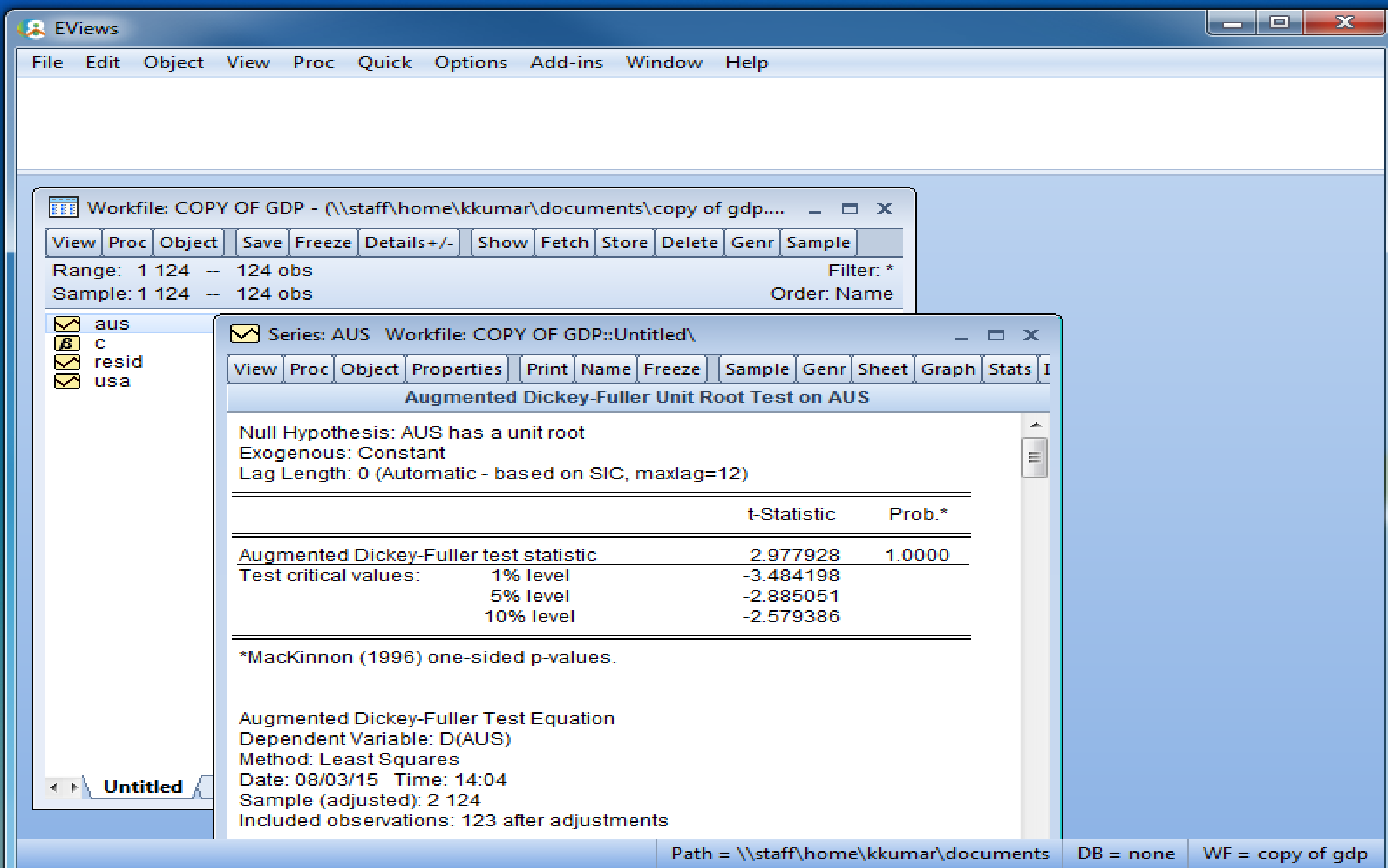
Unit Root Test - ADF @ level: Example using E-Views (Kumar, 2015)

BAZE UNIVERSITY ABUJA



Unit Root Test - ADF @ level:

Continuation - E-Views output (Kumar, 2015)



EViews

File Edit Object View Proc Quick Options Add-ins Window Help

Workfile: COPY OF GDP - (\\staff\home\kkumar\documents\copy of gdp....

View Proc Object Save Freeze Details+/- Show Fetch Store Delete Genr Sample

Range: 1 124 -- 124 obs Filter: *

Sample: 1 124 -- 124 obs Order: Name

aus
c
resid
usa

Untitled

Series: AUS Workfile: COPY OF GDP::Untitled\

View Proc Object Properties Print Name Freeze Sample Genr Sheet Graph Stats I

Augmented Dickey-Fuller Unit Root Test on AUS

Null Hypothesis: AUS has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		2.977928	1.0000
Test critical values:	1% level	-3.484198	
	5% level	-2.885051	
	10% level	-2.579386	

*MacKinnon (1996) one-sided p-values.

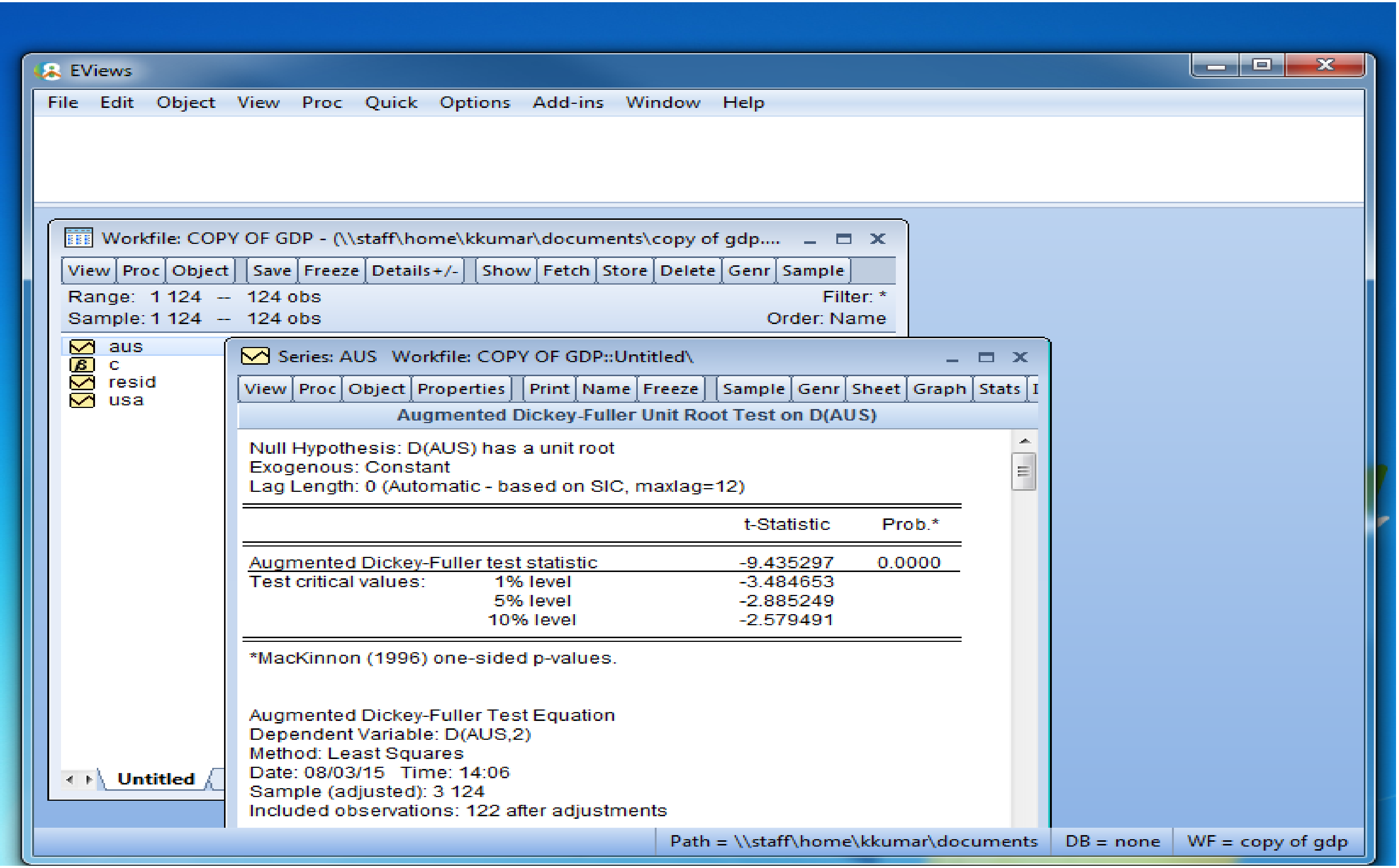
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(AUS)
Method: Least Squares
Date: 08/03/15 Time: 14:04
Sample (adjusted): 2 124
Included observations: 123 after adjustments

Path = \\staff\home\kkumar\documents DB = none WF = copy of gdp

Unit Root Test - ADF @ 1st difference:

Continuation - E-Views output (Kumar, 2015)

BAZE
UNIVERSITY
ABUJA



Stationarity test continuation (Kumar, 2015)

BAZE UNIVERSITY ABUJA

Regression model	1%	5%	10%
(1) $y_t = \beta x_t + e_t$	–3.39	–2.76	–2.45
(2) $y_t = \beta_1 + \beta_2 x_t + e_t$	–3.96	–3.37	–3.07
(3) $y_t = \beta_1 + \delta t + \beta_2 x_t + e_t$	–3.98	–3.42	–3.13

Note: These critical values are taken from J. Hamilton (1994), *Time Series Analysis*, Princeton University Press, p. 766.

Some information:
The Phillips-Perron (PP) unit root tests

BAZE UNIVERSITY ABUJA

- The Phillips-Perron (PP) unit root tests developed by Phillips and Perron (1988) are similar to ADF tests.
- Kumar (2015) suggests that the PP unit root tests differ from the ADF tests mainly in complexity and how they deal with serial correlation and heteroskedasticity in the errors.
- Notably, where the ADF tests use a parametric autoregression to approximate the ARMA structure of the errors in the test regression, the PP tests ignore any serial correlation in the test regression.
- ***The PP tests usually give the same conclusions as the ADF tests*** (Kumar, 2015).

Next… how to conduct the Cointegration test?

BAZE UNIVERSITY ABUJA

?

Co-integration test…

- Conventionally, non-stationary time-series variables should not be used in regression models to avoid the problem of spurious regression
  – There is an exception to this rule when

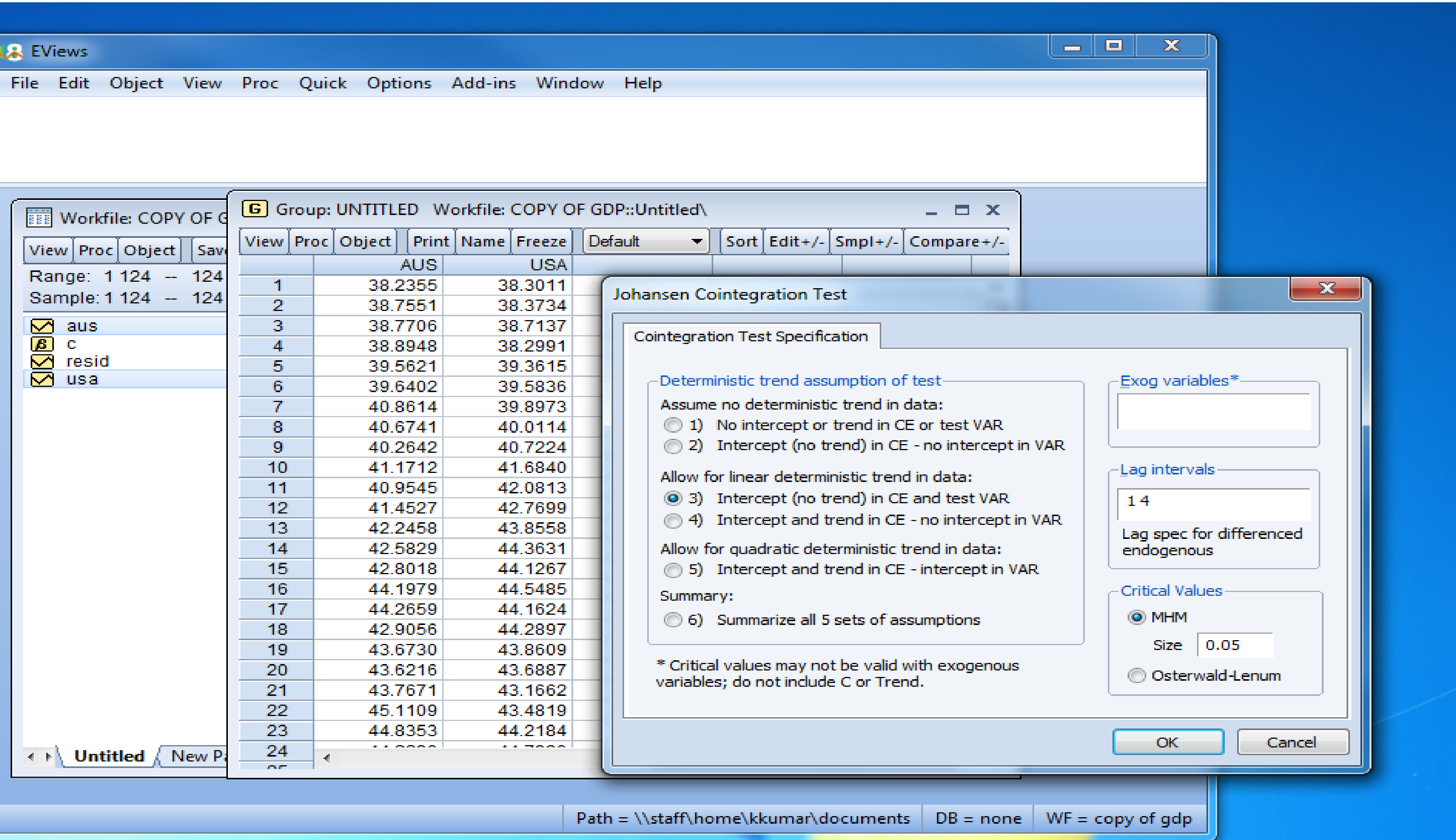
$e_t = y_t - \beta_1 - \beta_2 x_t$ is a stationary I(0) process

- In this case y_t and x_t are said to be **co-integrated,** i.e. y_t and x_t share similar stochastic trends, and, since the difference e_t is stationary, they never diverge too far from each other (Kumar, 2015).

Johansen Cointegration Test:

E-Views window(Kumar, 2015) – PRESS OK!

BAZE UNIVERSITY ABUJA



Johansen Cointegration Test:

E-Views output (Kumar, 2015) – At most 1 cointegration observed

BAZE UNIVERSITY ABUJA

EViews

File Edit Object View Proc Quick Options Add-ins Window Help

Workfile: COPY OF GDP - (\\staff\

View Proc Object Save Freeze Det

Range: 1 124 -- 124 obs
Sample: 1 124 -- 124 obs

aus
c
resid
usa

Untitled New Page

Group: UNTITLED Workfile: COPY OF GDP::Untitled\

View Proc Object Print Name Freeze Sample Sheet Stats Spec

Johansen Cointegration Test

Date: 08/03/15 Time: 14:13
Sample (adjusted): 6 124
Included observations: 119 after adjustments
Trend assumption: Linear deterministic trend
Series: AUS USA
Lags interval (in first differences): 1 to 4

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None	0.073266	13.98836	15.49471	0.0833
At most 1 *	0.040613	4.933864	3.841466	0.0263

Trace test indicates no cointegration at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None	0.073266	9.054500	14.26460	0.2816

Path = \\staff\home\kkumar\documents DB = none WF = copy of gdp

Next… how to conduct further tests using

VECM?

BAZE UNIVERSITY ABUJA

?

Recall:
Order of integration of the variables….

BAZE UNIVERSITY ABUJA

- Note that all variables have to be of the same order of integration; the following are possible cases (Hatemi, 2004):

1. *All the variables are I(0) (stationary):* one is in the standard case, i.e. a VAR in level.
2. All the variables are I(*d*) (non-stationary) with $d > 0$
3. **The variables are co-integrated: the error correction term has to be included in the VAR. The model becomes a Vector Error Correction Model (VECM).**
4. The variables are not *co-integrated* : the variables have first to be differenced *d* times and one has a VAR in difference.

Granger causality – VAR Equation (Kumar, 2015)

BAZE UNIVERSITY ABUJA

- In the VAR equation, the example we proposed above implies a lower triangular coefficient matrix:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \phi_{11}^1 & 0 \\ \phi_{21}^1 & \phi_{22}^1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \cdots + \begin{bmatrix} \phi_{11}^p & 0 \\ \phi_{21}^p & \phi_{22}^p \end{bmatrix} \begin{bmatrix} x_{t-p} \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

Or if we use MA representations,

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \Psi_{11}(B) & 0 \\ \Psi_{21}(B) & \Psi_{22}(B) \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

$$where\ \Psi_{ij}(B) = \phi_{ij}^0 + \phi_{ij}^1 B + \phi_{ij}^2 B^2 + \cdots, \phi_{11}^0 = \phi_{22}^0 = 1, \phi_{21}^0 = 0.$$

General procedure for testing Granger Causality… (Kumar, 2015)

BAZE UNIVERSITY ABUJA

1) Check that both series are stationary in mean, variance and covariance (if necessary transform the data via logs, differences to ensure this)
2) Estimate AR(*p*) models for each series, where *p* is large enough to ensure white noise residuals. F tests and other criteria can be used to establish the maximum lag *p* that is needed.
3) Re-estimate both model, now including all the lags of the other variable
4) Use F tests to determine whether, after controlling for past *Y*, past values of *X* can improve forecasts *Y* (and vice versa)

Testing Granger causality - continuation: (Kumar, 2015)

BAZE UNIVERSITY ABUJA

i. The simplest test is to estimate the regression which is based on

$$x_t = c_1 + \sum_{i=0}^{p} \alpha_i x_{t-i} + \sum_{j=1}^{p} \beta_j y_{t-j} + u_t$$

using **OLS** and then conduct a *F*-test of the null hypothesis

$$H_0 : \beta_1 = \beta_2 = . . . = \beta_p = 0.$$

Testing Granger causality continuation…

BAZE UNIVERSITY ABUJA

ii. Run the following regression, and calculate RSS (full model)

$$x_t = c_1 + \sum_{i=0}^{p} \alpha_i x_{t-i} + \sum_{j=1}^{p} \beta_j y_{t-j} + u_t$$

iii. Run the following limited regression, and calculate RSS (Restricted model).

$$\boldsymbol{x_t = c_1 + \sum_{i=0}^{p} \alpha_i x_{t-i} + u_t}$$

Testing Granger causality continuation…

BAZE UNIVERSITY ABUJA

iv. Do the following *F*-test using RSS obtained from stages 2 and 3:

$$F = [\{(n-k)/q\}.\{(RSSrestricted - RSSfull)/RSSfull\}]$$

n: number of observations

k: number of parameters from full model

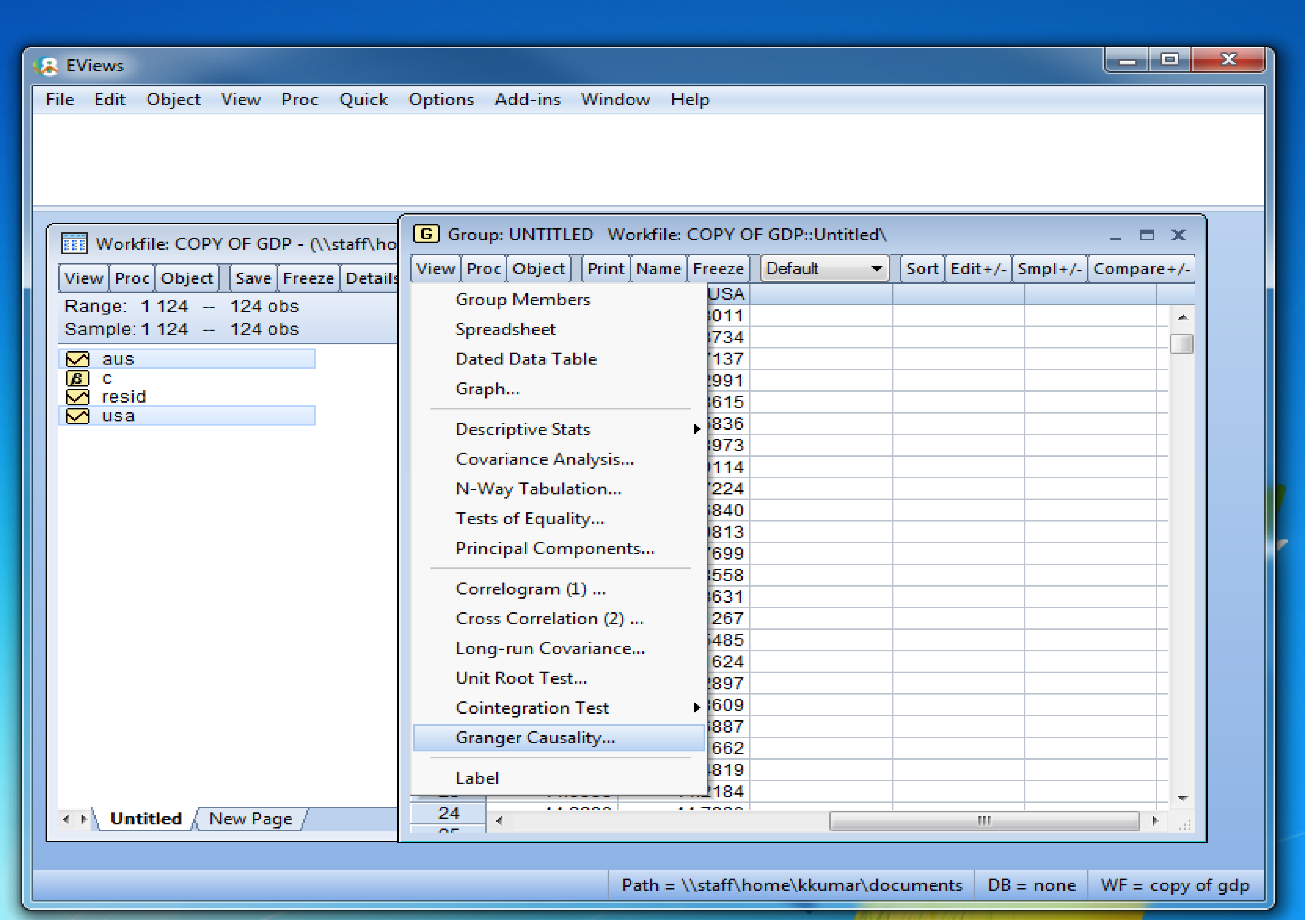
q: number of parameters from restricted model

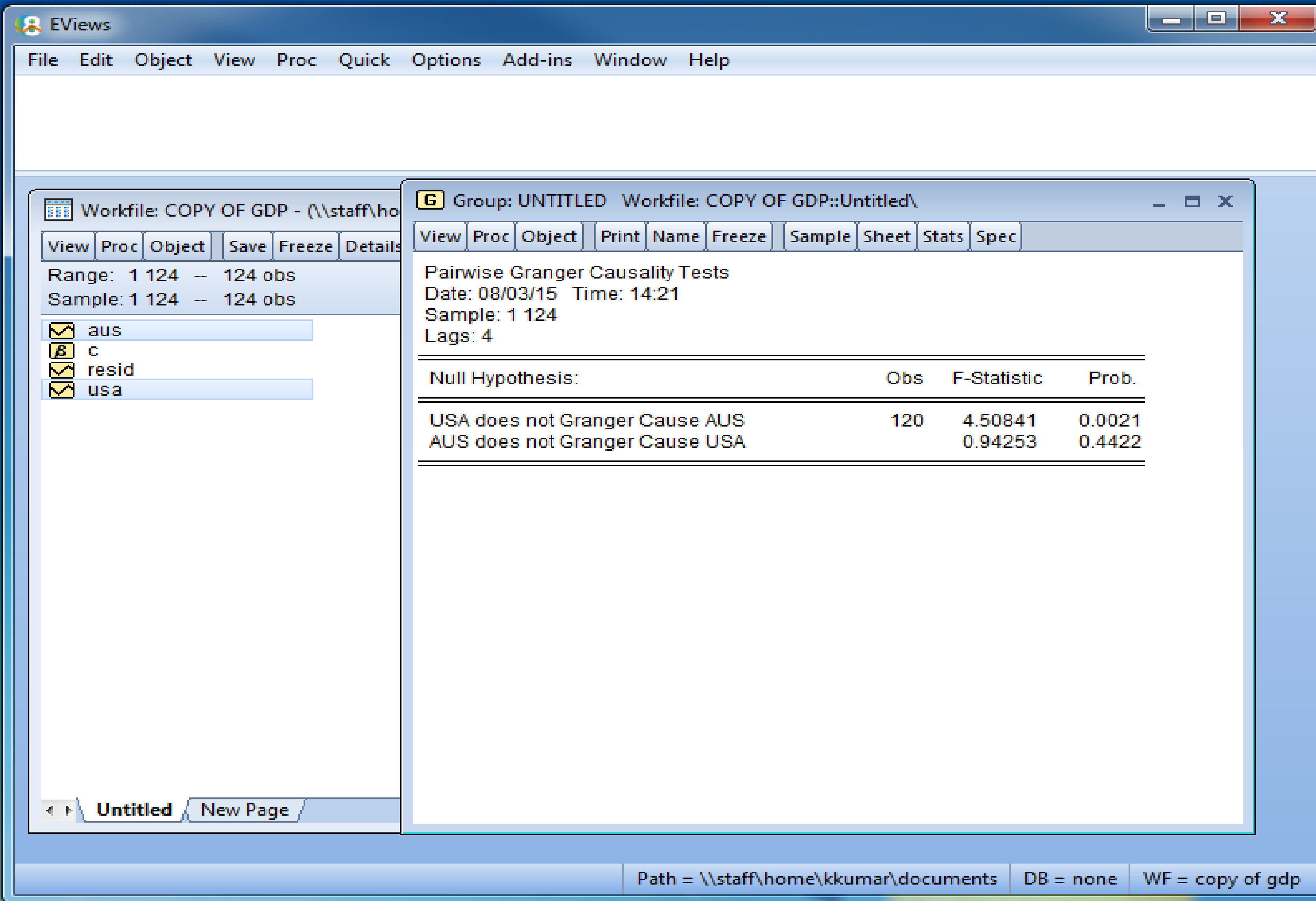
TESTING GRANGER CAUSALITY
(Kumar, 2015)

BAZE UNIVERSITY ABUJA

v. If H_0 rejected, then *X* causes *Y*.

- Granger test can be used in investigating whether or not *Y* causes *X*.


EViews
File Edit Object View Proc Quick Options Add-ins Window Help
Workfile: COPY OF GDP - (\\staff\ho
View Proc Object Save Freeze Details
Range: 1 124 -- 124 obs
Sample: 1 124 -- 124 obs
aus
c
resid
usa
Untitled
New Page
Group: UNTITLED Workfile: COPY OF GDP::Untitled\
View Proc Object Print Name Freeze Default Sort Edit+/- Smpl+/- Compare+/-
Group Members
Spreadsheet
Dated Data Table
Graph...
Descriptive Stats
Covariance Analysis...
N-Way Tabulation...
Tests of Equality...
Principal Components...
Correlogram (1) ...
Cross Correlation (2) ...
Long-run Covariance...
Unit Root Test...
Cointegration Test
Granger Causality...
Label
USA
Path = \\staff\home\kkumar\documents
DB = none
WF = copy of gdp


EViews
File Edit Object View Proc Quick Options Add-ins Window Help
Workfile: COPY OF GDP - (\\staff\ho
View Proc Object Save Freeze Details
Range: 1 124 -- 124 obs
Sample: 1 124 -- 124 obs
aus
c
resid
usa
Untitled
New Page
Group: UNTITLED Workfile: COPY OF GDP::Untitled\
View Proc Object Print Name Freeze Sample Sheet Stats Spec
Pairwise Granger Causality Tests
Date: 08/03/15 Time: 14:21
Sample: 1 124
Lags: 4
Null Hypothesis: Obs F-Statistic Prob.
USA does not Granger Cause AUS 120 4.50841 0.0021
AUS does not Granger Cause USA 0.94253 0.4422
Path = \\staff\home\kkumar\documents
DB = none
WF = copy of gdp

Next Session…

- Review of some practical applications

Session 2

Exploring some applications using EViews econometric software

Note: some helpful technology hints (EViews software)

BAZE UNIVERSITY ABUJA

For what?	Where to go on the EViews package
i. Graph Options	Select VIEW
ii. Unit Root Test	Select VIEW
iii. Cointegration Test	Select VIEW
iv. Vector Error Correction	Select PROC
v. Granger Causality	Select VIEW

Beyond analytics:
Some important workplace skills…

BAZE UNIVERSITY ABUJA

✓Written & oral communication

✓Problem-solving abilities – analytical reasoning

✓Creative thinking

✓Interpersonal skills (team work)

✓Time management

✓General professionalism, personal effectiveness, and work ethics

Reviewing cases in Financial Econometrics

BAZE UNIVERSITY ABUJA

?

Recall: some helpful technology hints (EViews software)

BAZE UNIVERSITY ABUJA

For what?	Where to go on the EViews package
i. Graphs	Select VIEW
ii. Unit Root test	Select VIEW
iii. Cointegration test	Select VIEW
iv. Vector Error Correction	Select PROC
v. Granger Causality	Select VIEW

Granger causality analysis

Now, let's test these pairs of time series…

BAZE UNIVERSITY ABUJA

1. Econodata file_1: DGR & GDP growth in Nigeria
2. Econodata file_2: Agriculture & GDP in Nigeria
3. Econodata file_3: Crude oil price and growth in Nigeria
4. Econodata file_4: Infosys Ltd & NSE Nifty - India
5. Econodata file_5: Agriculture & GDP in India
6. Econodata file_6: India economic growth & Global growth

Econodata file_1:
DGR & GDP growth in Nigeria

BAZE UNIVERSITY ABUJA

Situation analysis of Debt-to-GDP Ratio (DGR) and GDP in Nigeria?

Causality analysis
Experimenting with Case Nigeria…

BAZE UNIVERSITY ABUJA

- **CONTEXT:** The assumption that the developing countries can borrow with expectation of economic growth is open to question.
- It's controversial – Government needs to borrow to fulfil huge developmental goals for their citizens, but if government becomes a dominant debtor in a financial system, there is concern that the private sector may become 'growth at the end.
- Besides the economic implications and associated debate on the subject, the degree of stability or volatility of government's fiscal policies such as Debt-to-GDP ratio (DGR) will have remarkable influence on business performance because ***companies are not immuned from the macroeconomic environment in which they operate***.

Formulate your hypotheses

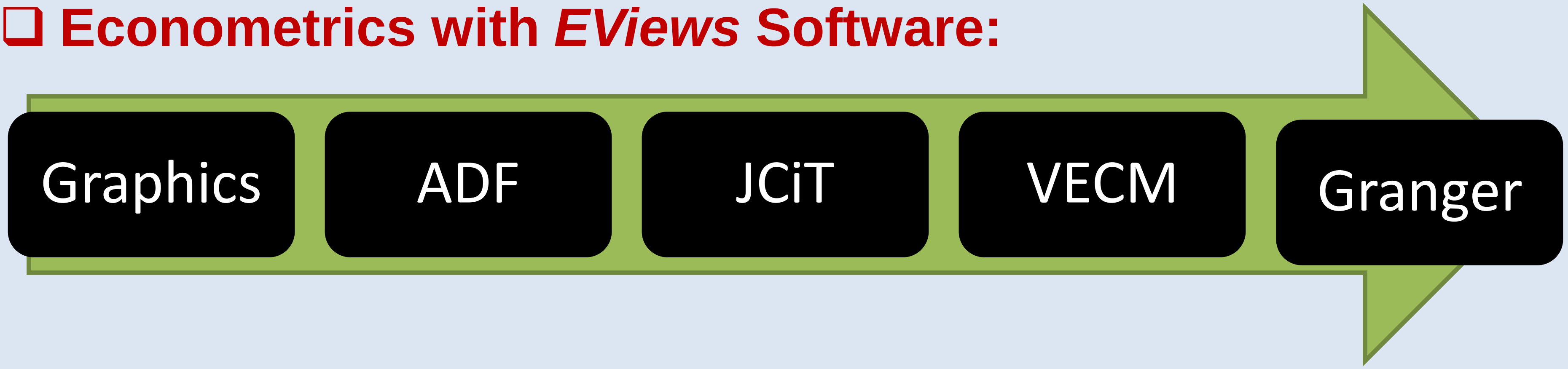
BAZE UNIVERSITY ABUJA

- $H_{1:}$ GDP growth rate has a unit root.
- $H_{2:}$ DGR has a unit root.
- $H_{3:}$ There is no co-integration between GDP and DGR.
- $H_{4:}$ GDP growth rate does not Granger-cause DGR.
- $H_{5:}$ DGR does not Granger-cause GDP growth rate.

DGR-GDP growth causality test:
An application of the Granger methodology…

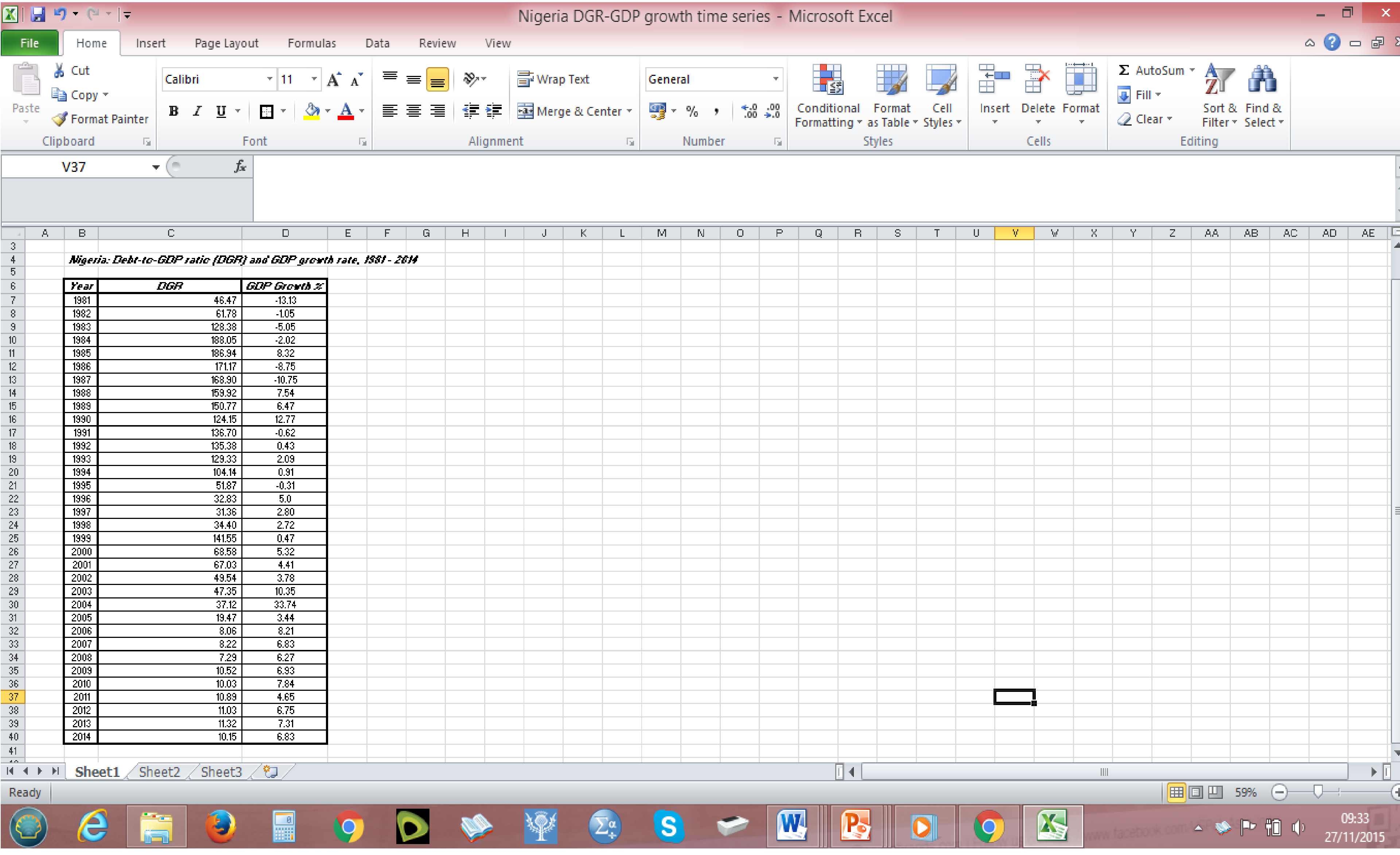
BAZE UNIVERSITY ABUJA

- ***Data:*** (i) ***GDP growth rate*** was used as the proxy for economic growth- time series from 1981 to 2014; i.e. 34 years of secondary data sourced primarily from The World Bank and the Central bank of Nigeria (CBN). (ii) **Debt** means 'total debt stock', i.e. it includes long-term and short-term domestic and foreign liabilities.
- **Econometrics with *EViews* Software:**



Next…

Upload your data for analysis



Nigeria DGR-GDP growth time series - Microsoft Excel

Nigeria: Debt-to-GDP ratio (DGR) and GDP growth rate, 1981 - 2014

Year	DGR	GDP Growth %
1981	46.47	-13.13
1982	61.78	-1.05
1983	128.38	-5.05
1984	188.05	-2.02
1985	186.94	8.32
1986	171.17	-8.75
1987	168.90	-10.75
1988	159.92	7.54
1989	150.77	6.47
1990	124.15	12.77
1991	136.70	-0.62
1992	135.38	0.43
1993	129.33	2.09
1994	104.14	0.91
1995	51.87	-0.31
1996	32.83	5.0
1997	31.36	2.80
1998	34.40	2.72
1999	141.55	0.47
2000	68.58	5.32
2001	67.03	4.41
2002	49.54	3.78
2003	47.35	10.35
2004	37.12	33.74
2005	19.47	3.44
2006	8.06	8.21
2007	8.22	6.83
2008	7.29	6.27
2009	10.52	6.93
2010	10.03	7.84
2011	10.89	4.65
2012	11.03	6.75
2013	11.32	7.31
2014	10.15	6.83

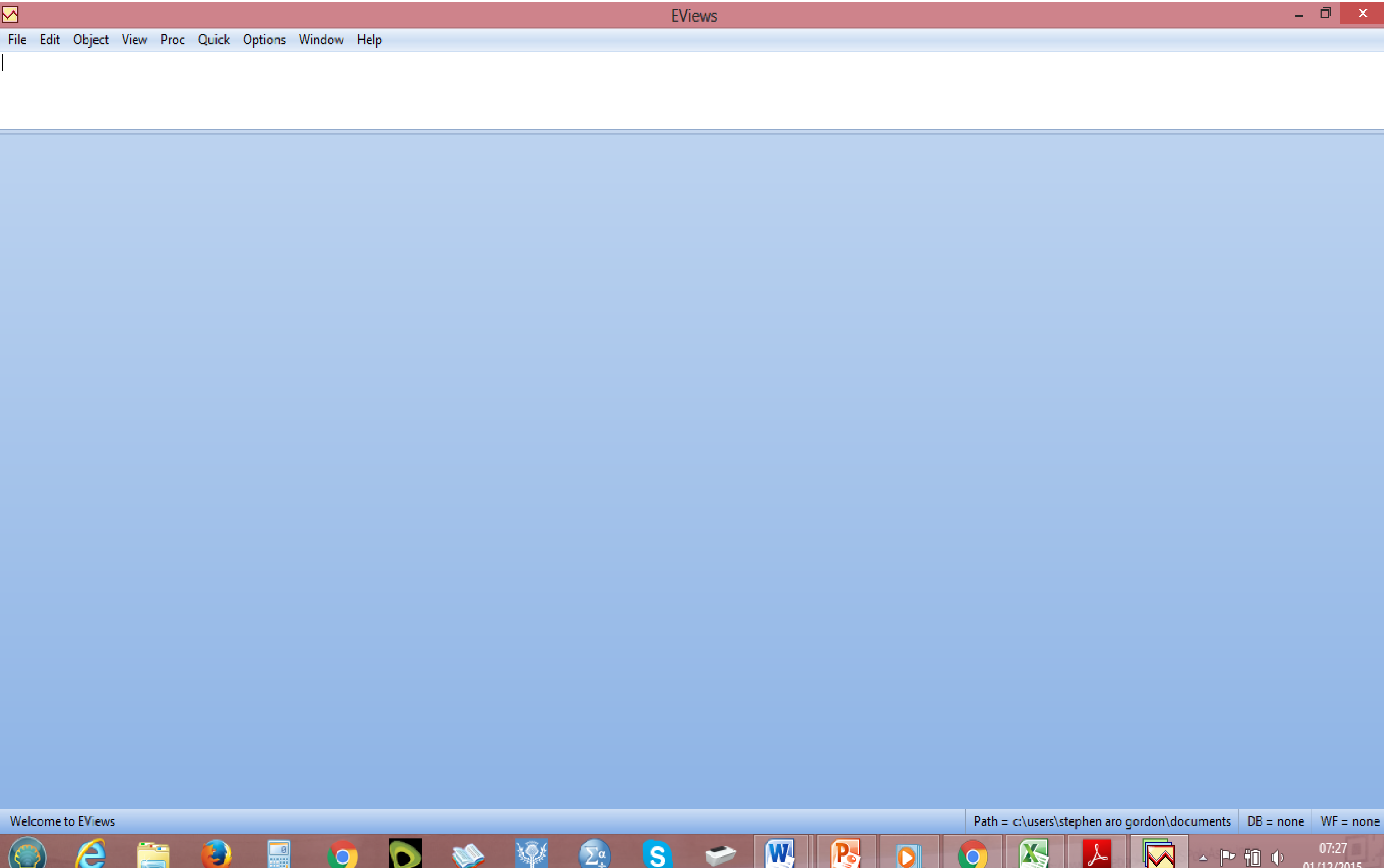
Sheet1 Sheet2 Sheet3

First, upload your data into the EViews …..

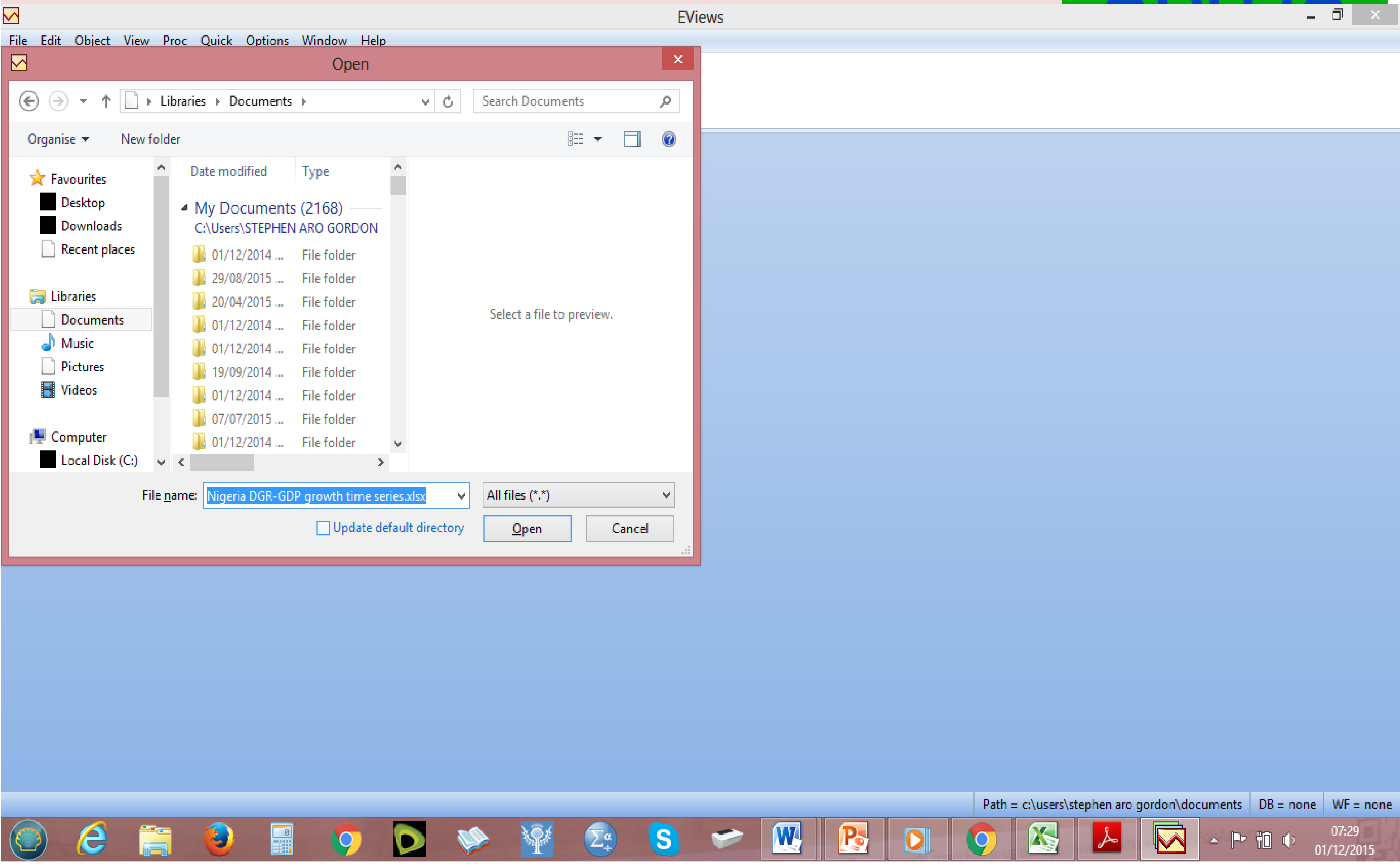
BAZE UNIVERSITY ABUJA

- Click-open your *EViews* – FILE-OPEN-FOREIGN DATA AS WORK FILE-locate your file in MS Excel-OPEN-NEXT-FINISH

You're ready!
Your EViews now opened…



Next, open your time series data



Next,

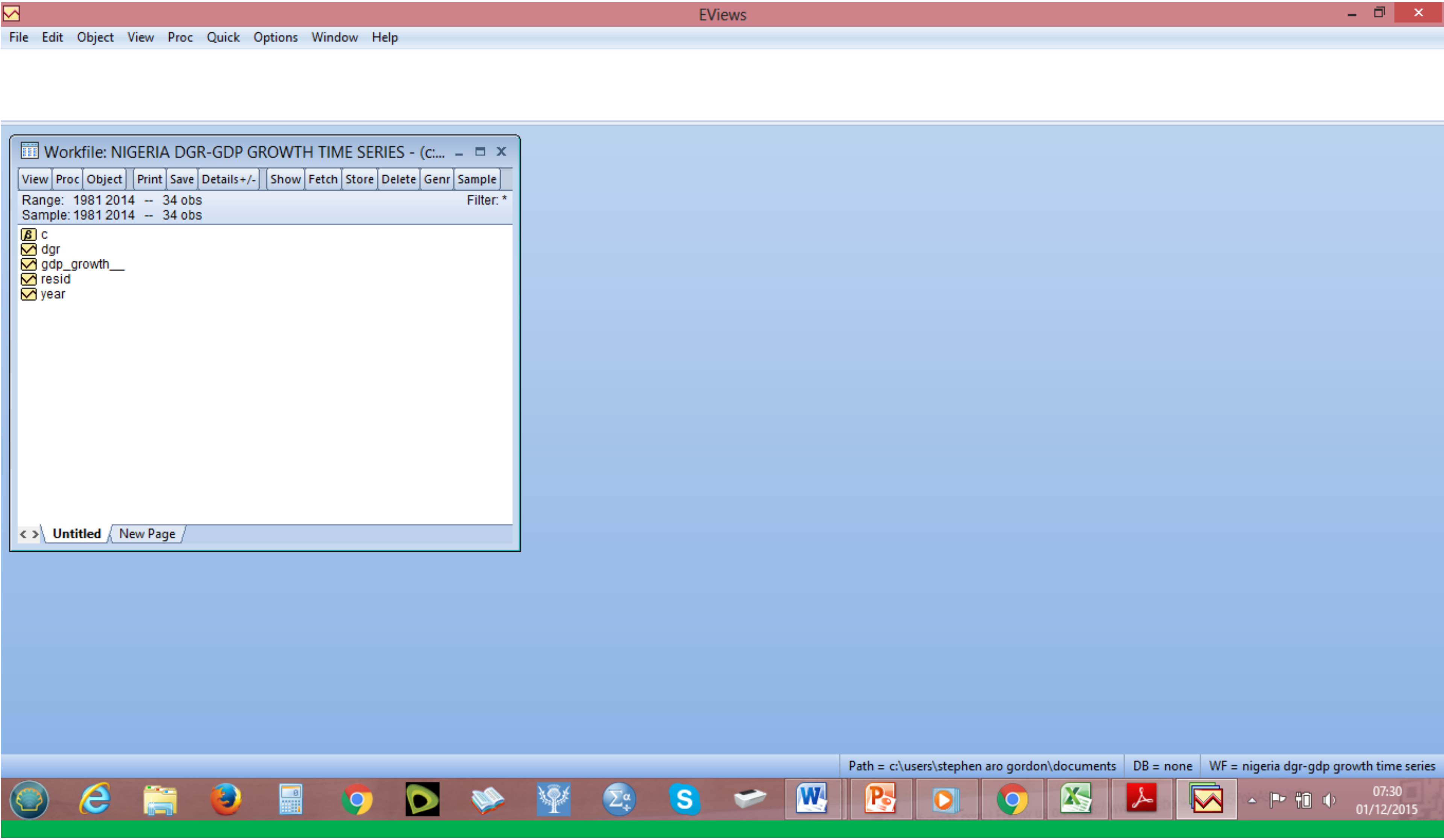
Obtain the time series' <u>graphical visuals</u>

- Highlight the <u>two</u> time series (DGR&GDP growth)-<u>**VIEW**</u>-OPEN SELECTED-ONE WINDOW-OPEN GROUP (to display the two time series)-<u>**VIEW**</u>(again)-GRAPH-GRAPH OPTIONS-LINE&SYMBOL-PRESS OK.

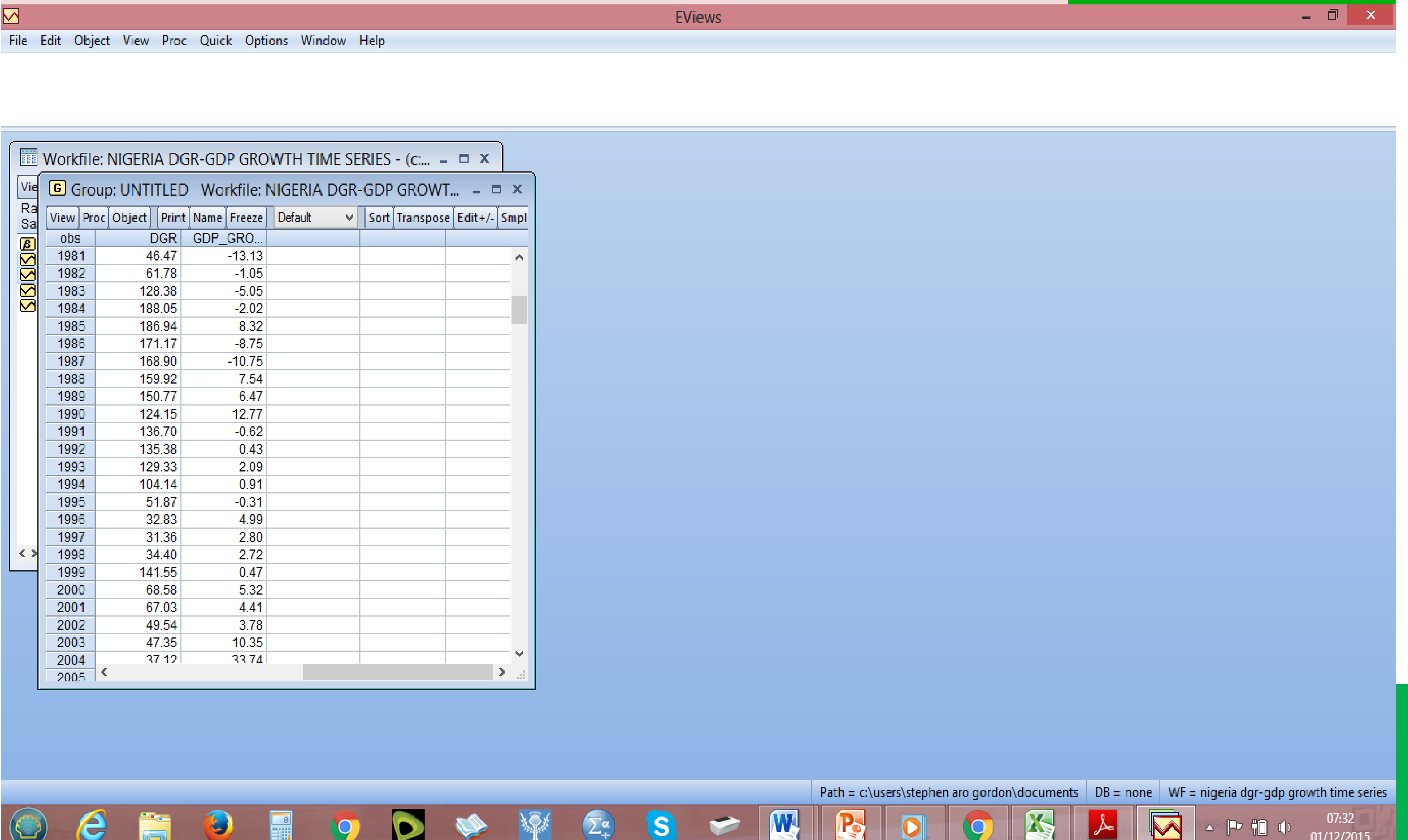
EViews graphic visualization example

Remember to highlight / select the two time series…

BAZE UNIVERSITY ABUJA

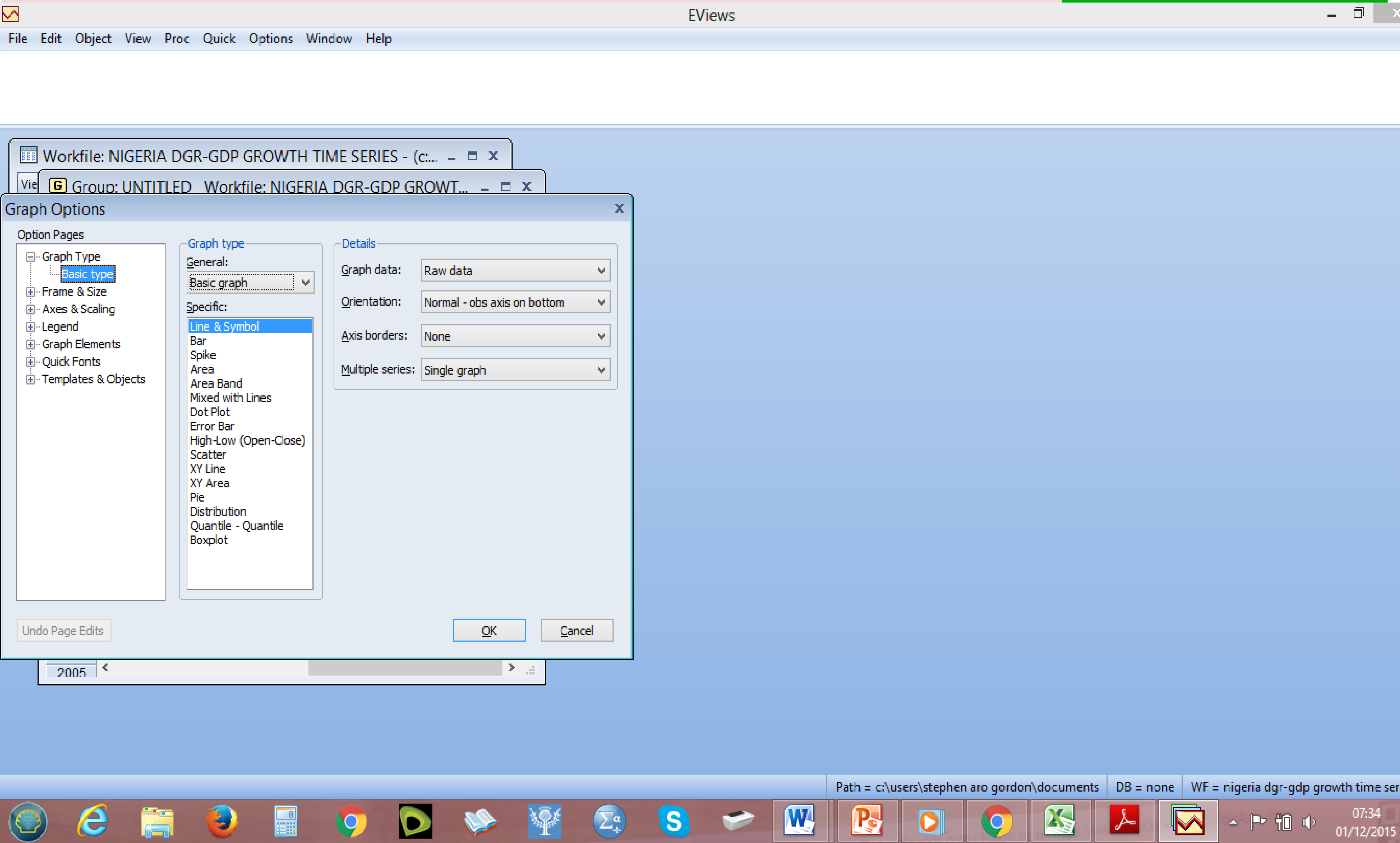


The two time series displayed…



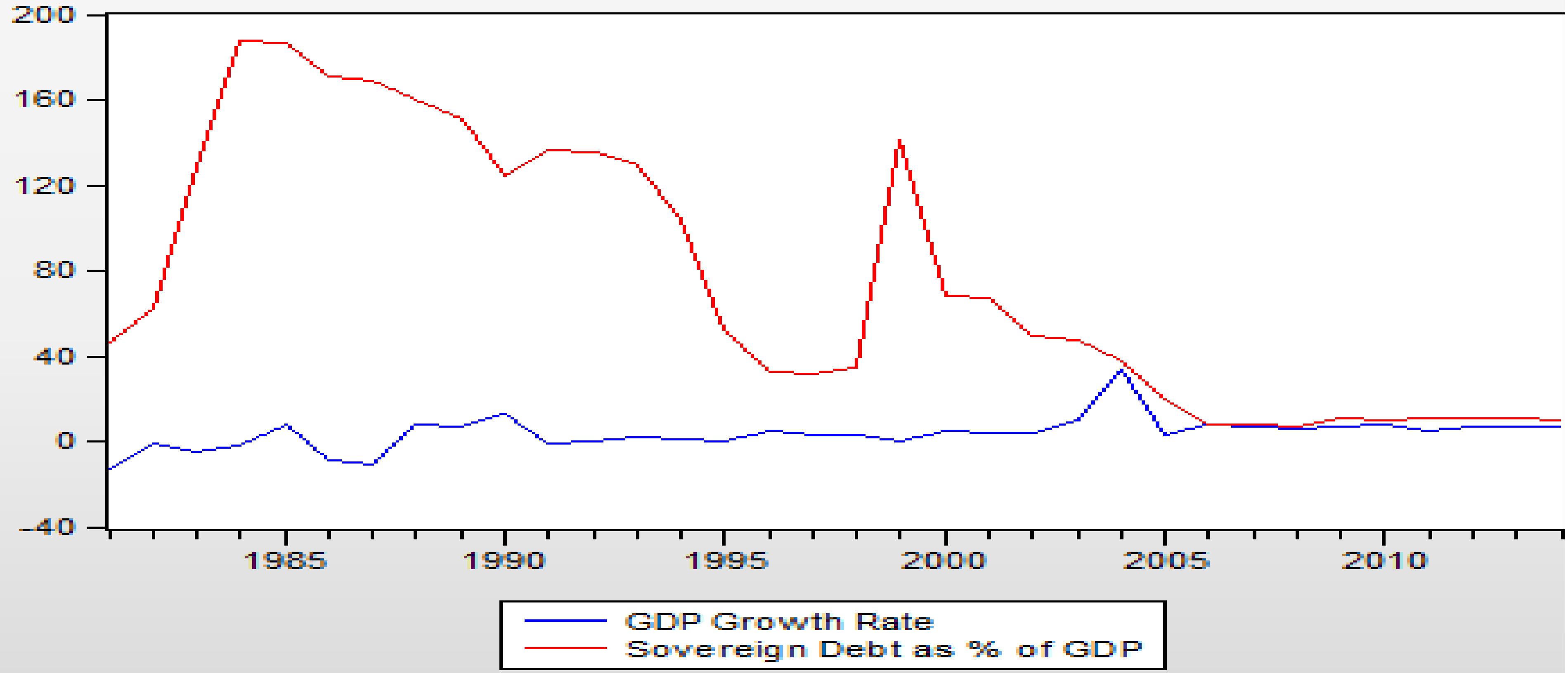
obs	DGR	GDP_GRO...
1981	46.47	-13.13
1982	61.78	-1.05
1983	128.38	-5.05
1984	188.05	-2.02
1985	186.94	8.32
1986	171.17	-8.75
1987	168.90	-10.75
1988	159.92	7.54
1989	150.77	6.47
1990	124.15	12.77
1991	136.70	-0.62
1992	135.38	0.43
1993	129.33	2.09
1994	104.14	0.91
1995	51.87	-0.31
1996	32.83	4.99
1997	31.36	2.80
1998	34.40	2.72
1999	141.55	0.47
2000	68.58	5.32
2001	67.03	4.41
2002	49.54	3.78
2003	47.35	10.35
2004	37.12	33.74
2005		

Selecting your graph options…



Nigeria DGR-GDP growth time series

EViews graphic output



***Next, perform the Unit Root Test for each series i.e. individually* – let's start with DGR…**

BAZE UNIVERSITY ABUJA

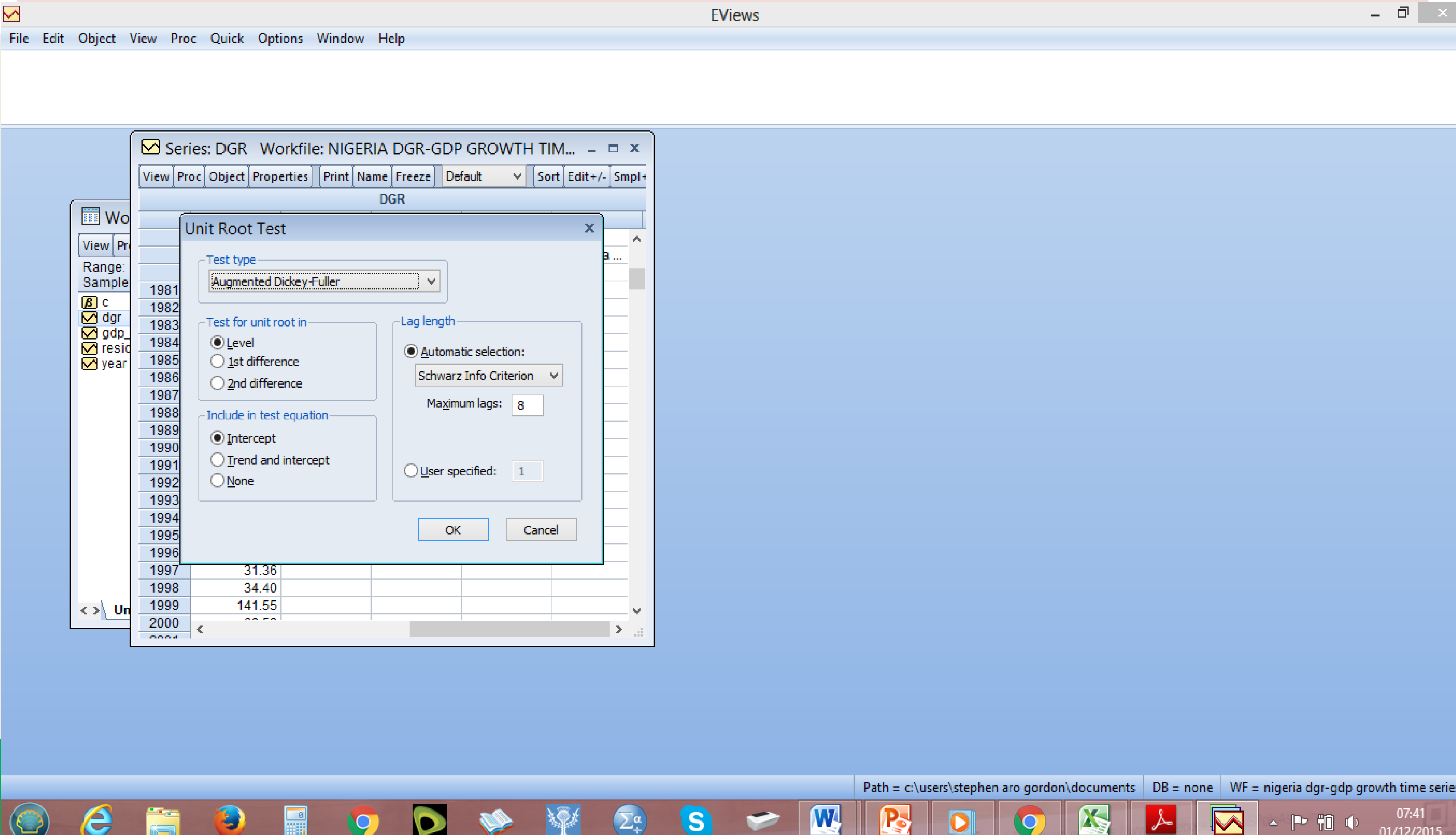
1. AT LEVEL

- HIGHLIGHT/SELECT THE TIME SERIES (DGR/GDP GROWTH-**VIEW-**OPEN SELECTED-**VIEW** (again)-UNIT ROOT TEST-AUGMENTED DICKEY-FULLER-LEVEL-PRESS OK

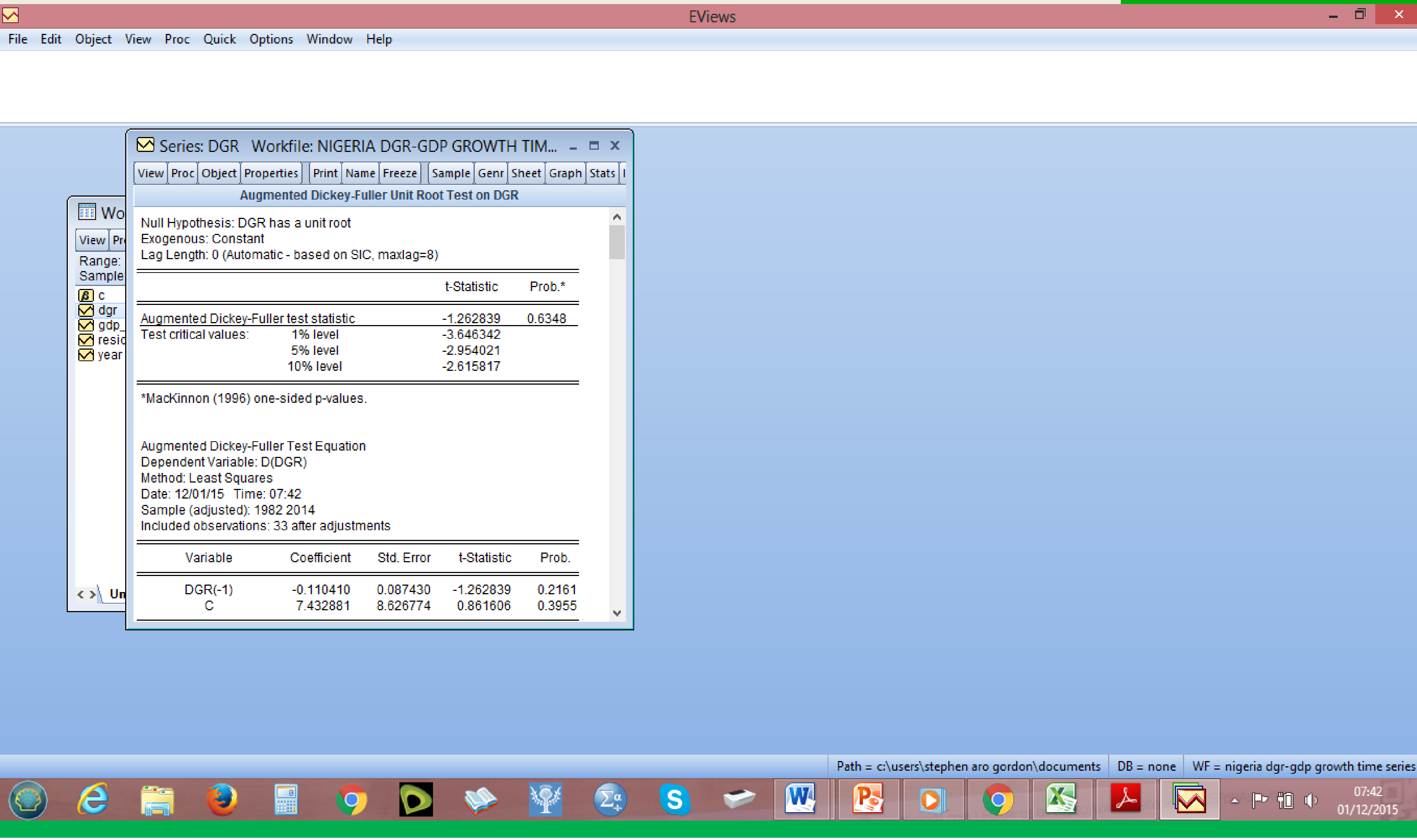
2. FIRST DIFFERENCE

- HIGHLIGHT/SELECT THE TIME SERIES (DGR/GDP GROWTH-**VIEW**-OPEN SELECTED-**VIEW** (again)-UNIT ROOT TEST-AUGMENTED DICKEY-FULLER-**1ST DIFFERENCE**-PRESS OK ***(check for stationarity: p≈0)***

EViews window - at level (default)



EViews window – URT output for DGR <u>at level</u>



Augmented Dickey-Fuller Unit Root Test on DGR

Null Hypothesis: DGR has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=8)

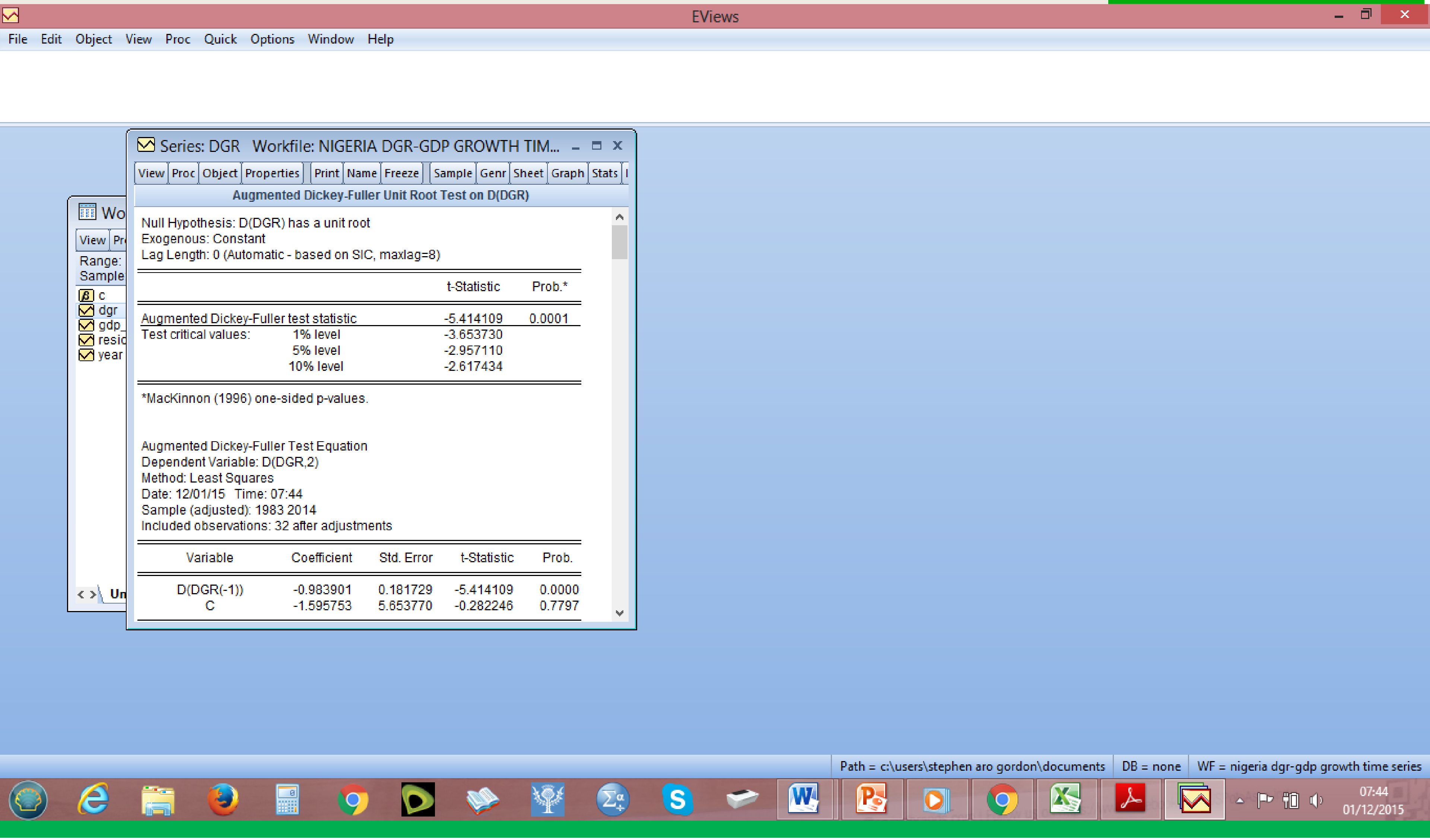
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.262839	0.6348
Test critical values:	1% level	-3.646342	
	5% level	-2.954021	
	10% level	-2.615817	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DGR)
Method: Least Squares
Date: 12/01/15 Time: 07:42
Sample (adjusted): 1982 2014
Included observations: 33 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DGR(-1)	-0.110410	0.087430	-1.262839	0.2161
C	7.432881	8.626774	0.861606	0.3955

EViews window – URT output for DGR at <u>first difference</u>



Series: DGR Workfile: NIGERIA DGR-GDP GROWTH TIM...

Augmented Dickey-Fuller Unit Root Test on D(DGR)

Null Hypothesis: D(DGR) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=8)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.414109	0.0001
Test critical values:	1% level	-3.653730	
	5% level	-2.957110	
	10% level	-2.617434	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DGR,2)
Method: Least Squares
Date: 12/01/15 Time: 07:44
Sample (adjusted): 1983 2014
Included observations: 32 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(DGR(-1))	-0.983901	0.181729	-5.414109	0.0000
C	-1.595753	5.653770	-0.282246	0.7797

DGR-GDP growth unit root test Summary of results

BAZE UNIVERSITY ABUJA

Particulars	DGR				GDP			
	t-statistic	Critical Value		P-value	t-statistic	Critical Value		P-value
At level	-1.2628	1%	-3.6463	0.6348	-4.6899	1%	-3.6463	0.0006
		5%	-2.9540			5%	-2.9540	
		10%	-2.6158			10%	-2.6158	
At first difference	-5.4141	1%	-3.6537	0.0001	-8.5980	1%	-3.6537	0.0000
		5%	-2.9571			5%	-2.9571	
		10%	-2.6174			10%	-2.6174	

A note before

Performing Cointegration test…

BAZE UNIVERSITY ABUJA

- In some situations, it may be desirable to determine the ***short-run*** dynamics or ***long-run*** dynamics exclusively.
- For instance, in technical analysis (as opposed to fundamental analysis) asset prices are believed to be long-run unpredictable but may have some short- or medium-run predictability

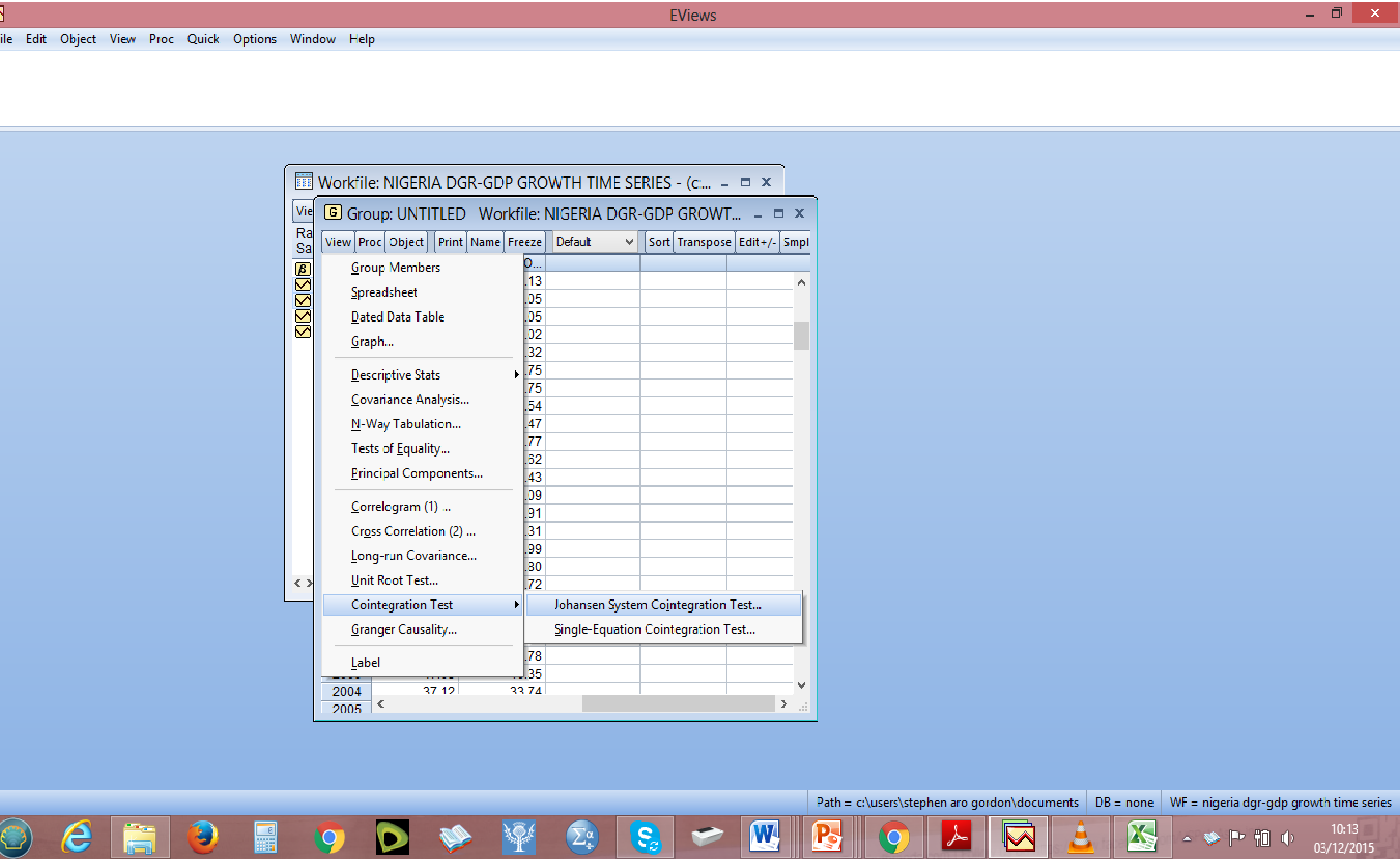
Next,

Perform Cointegration test…

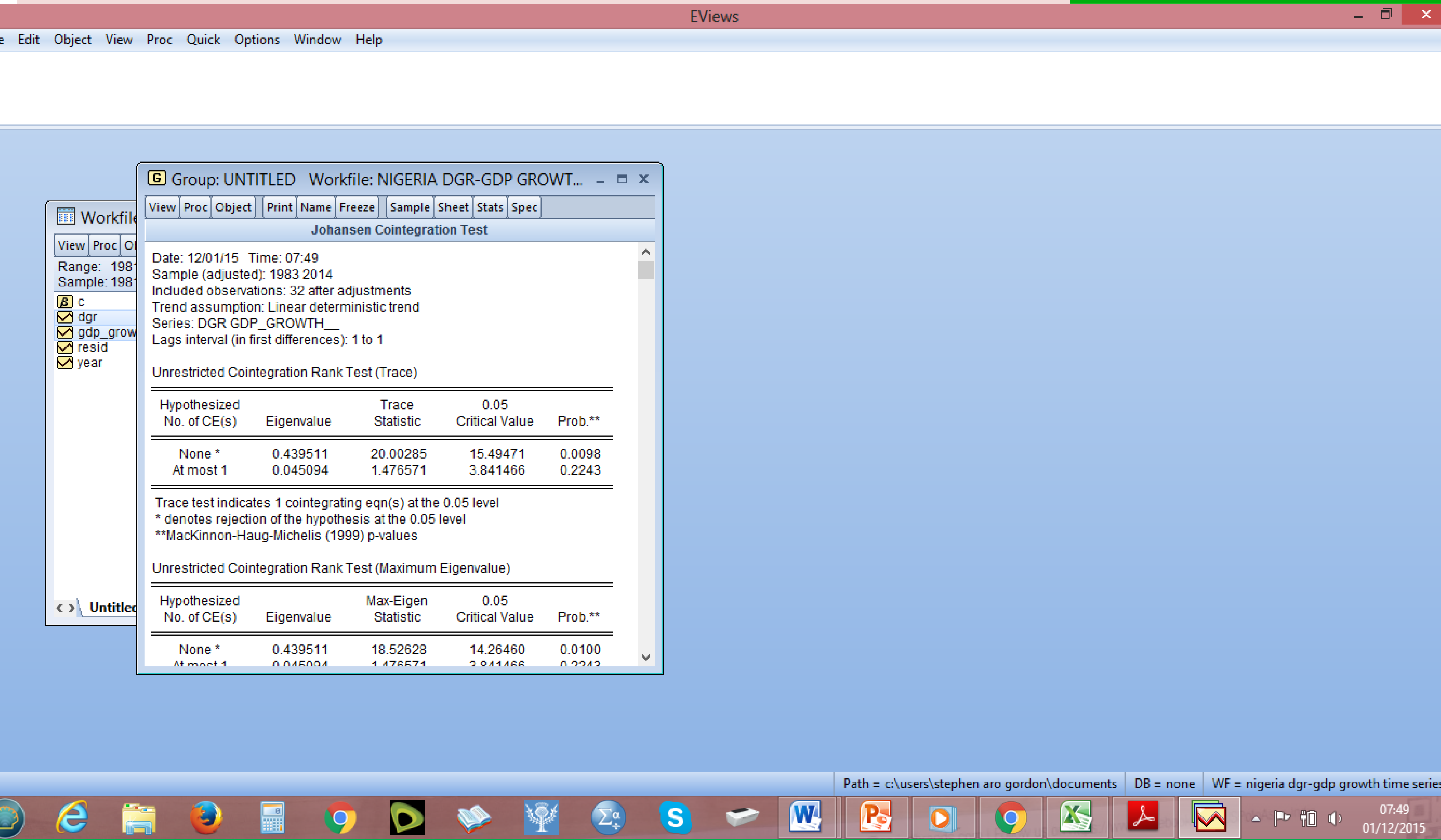
- Highlight the two time series (DGR&GDP growth)-**VIEW**-OPEN SELECTED-ONE WINDOW-OPEN GROUP (to display the two time series)-**VIEW**(again)-COINTEGRATION TEST-JOHANSEN SYSTEM COINTEGRATION-PRESS OK.

EViews windows now opened for *Johansen system Cointegration Test*

BAZE UNIVERSITY ABUJA



Johansen system Cointegration Test EViews Output



Johansen Cointegration Test

Date: 12/01/15 Time: 07:49
Sample (adjusted): 1983 2014
Included observations: 32 after adjustments
Trend assumption: Linear deterministic trend
Series: DGR GDP_GROWTH__
Lags interval (in first differences): 1 to 1

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.439511	20.00285	15.49471	0.0098
At most 1	0.045094	1.476571	3.841466	0.2243

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.439511	18.52628	14.26460	0.0100

Johansen system Cointegration Test

Summary results

BAZE UNIVERSITY ABUJA

Level	Eigen Value	Trace Statistic	Critical Value at 5%	P-values
H_o: r = 0 (none)*	0.4395	20.0028	15.4947	0.0098
H_1: r = 1 (at most 1)	0.0451	1.4766	3.8415	0.2243

Johansen system Cointegration Test

Summary results - continuation

- ***Scroll down*** to the bottom of the EVIEWS output where you have
  - ALPHA
  - COEFFICIENTS
- ***Observe:*** The estimated co-integrating coefficient for the GDP growth is as follows:

$$LGDP = -7.7045 - 0.1134DGR$$

[0.08]

- The t-statistic (standard error) of the co-integrating coefficient of DGR is given in the bracket suggesting significance at roughly 10%
- ***The coefficient for DGR is negative, at least in the short-run, which means that increase in DGR can be associated with decline in the country's economic growth.***

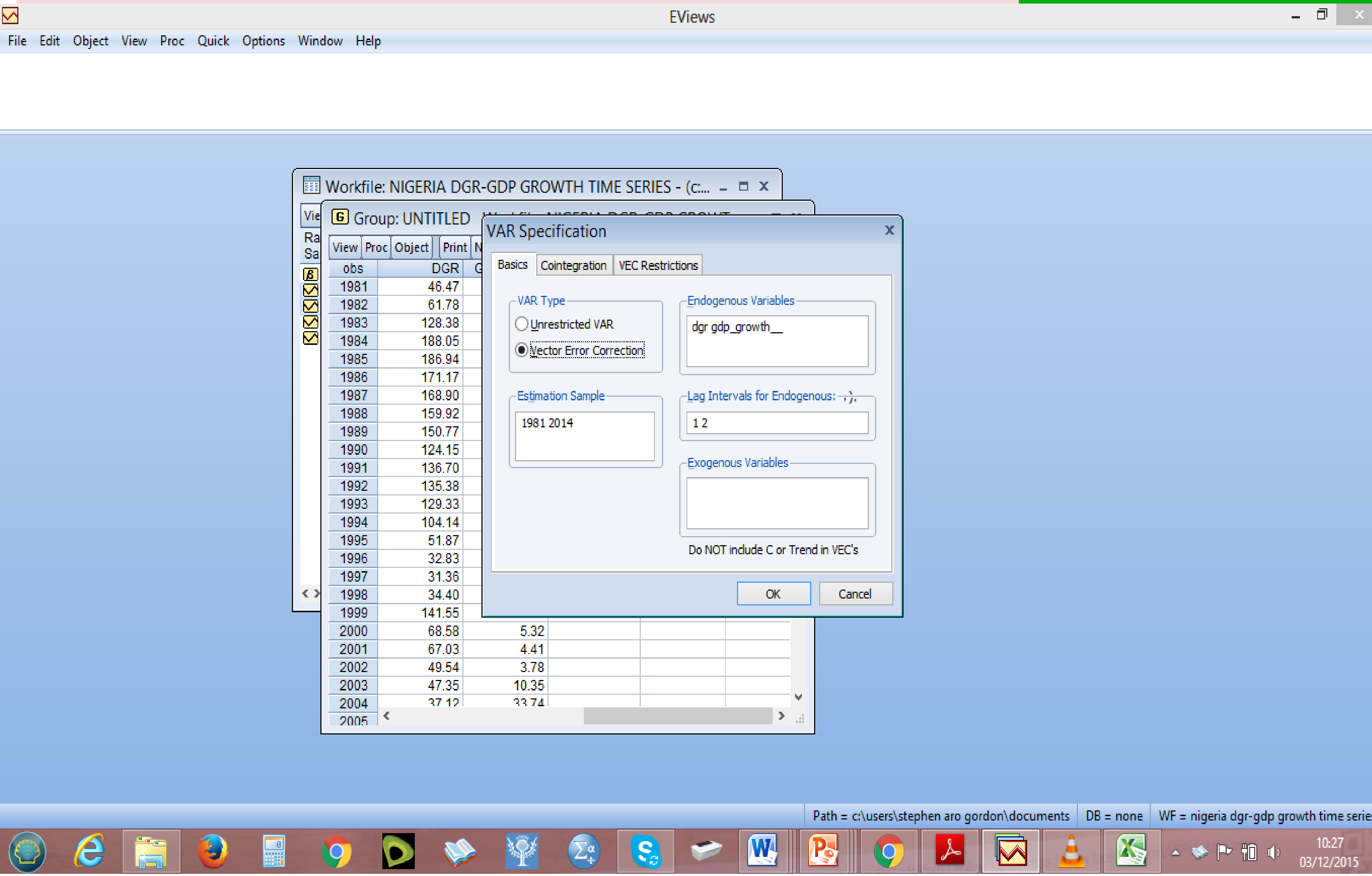
Next,

Perform Vector Error Correction test…

- Highlight the two time series (DGR&GDP growth)-**VIEW**-OPEN SELECTED-ONE WINDOW-**PROC**-MAKE VECTOR AUTOCORRECTION-VECTOR ERROR CORRECTION--PRESS OK.

EViews windows now opened for

Vector Error Correction Test



Vector Error Correction Test

EViews Output

BAZE UNIVERSITY ABUJA

Vector Error Correction Estimates
Date: 12/01/15 Time: 07:53
Sample (adjusted): 1984 2014
Included observations: 31 after adjustments
Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1	
DGR(-1)	1.000000	
GDP_GROWTH__(-1)	17.11125	
	(3.86788)	
	[4.42394]	
C	-154.8679	
Error Correction:	**D(DGR)**	**D(GDP_GR...**
CointEq1	-0.113375	-0.054897
	(0.08142)	(0.01971)
	[-1.39254]	[-2.78568]
D(DGR(-1))	-0.043143	0.038671
	(0.17706)	(0.04286)
	[-0.24367]	[0.90231]

VECM

SUMMARY RESULTS

Vector Error Correction Estimates

Cointegrating Eq:	CointEq1	
DGR(-1)	1	
GDP_GROWTH__(-1)	17.11125060891794 3.867876205049176 [4.42394]	
C	-154.8678962559113	
Error Correction:	**D(DGR)**	**D(GDP_GROWTH__)**
CointEq1	-0.1133751309235637 0.08141595564061155 [-1.39254]	-0.05489747606787678 0.01970701873928458 [-2.78568]
D(DGR(-1))	-0.04314304533812878 0.1770585545832603 [-0.24367]	0.03867107144165221 0.04285764658374208 [0.90231]
D(DGR(-2))	0.07814966611371196 0.1780213582101248 [0.43899]	-0.00662762780739984 0.04309069659178506 [-0.15381]
D(GDP_GROWTH__(-1))	1.000290430940833 1.120216515418953 [0.89294]	0.07595361121883884 0.271152351989411 [0.28011]
D(GDP_GROWTH__(-2))	0.4463946344693092 0.7912646447754125 [0.56415]	-0.04282801058257431 0.1915283934165866 [-0.22361]
C	-4.350778297955078 5.596055445934629	0.4455776708401139 1.354544925148543

VECM tests
SUMMARY RESULTS

BAZE UNIVERSITY ABUJA

- ***Note***: t-statistics in []
- ***Scroll through the output / carefully observe***: In all cases t < 1.96 (alpha)
- **Decision**: Null Hypothesis is accepted – this means that there may be no long-run cointegration between DGR and GDP growth rate.

Finally,

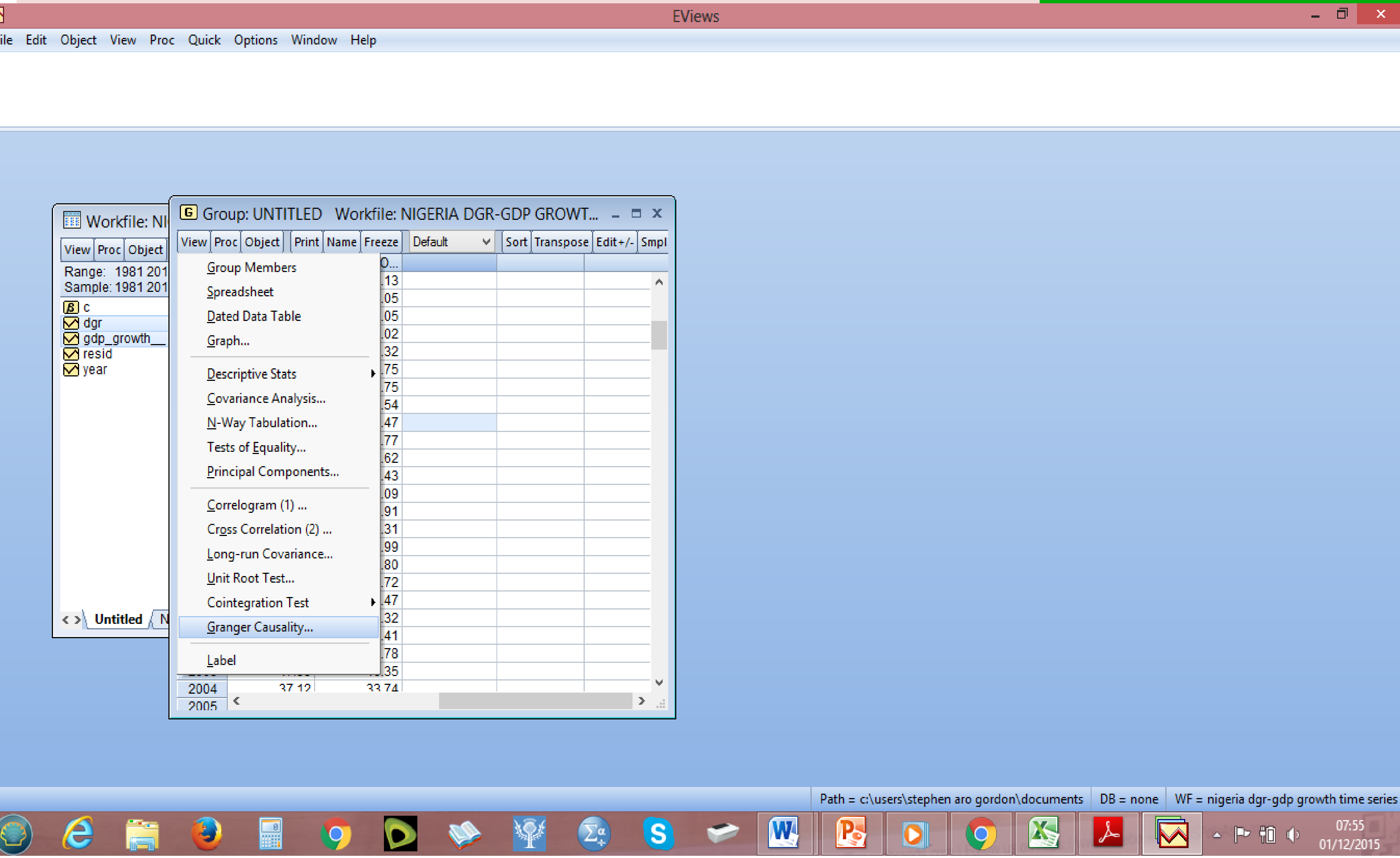
Perform your <u>Granger Causality test</u>…

BAZE UNIVERSITY ABUJA

- Highlight the <u>two</u> time series (DGR&GDP growth)-<u>**VIEW**</u>-OPEN SELECTED-ONE WINDOW-OPEN GROUP (to display the two time series)-<u>**VIEW**</u>(again)-GRANGER CAUSALITY-LAG SPECIFICATION(2)-PRESS OK.

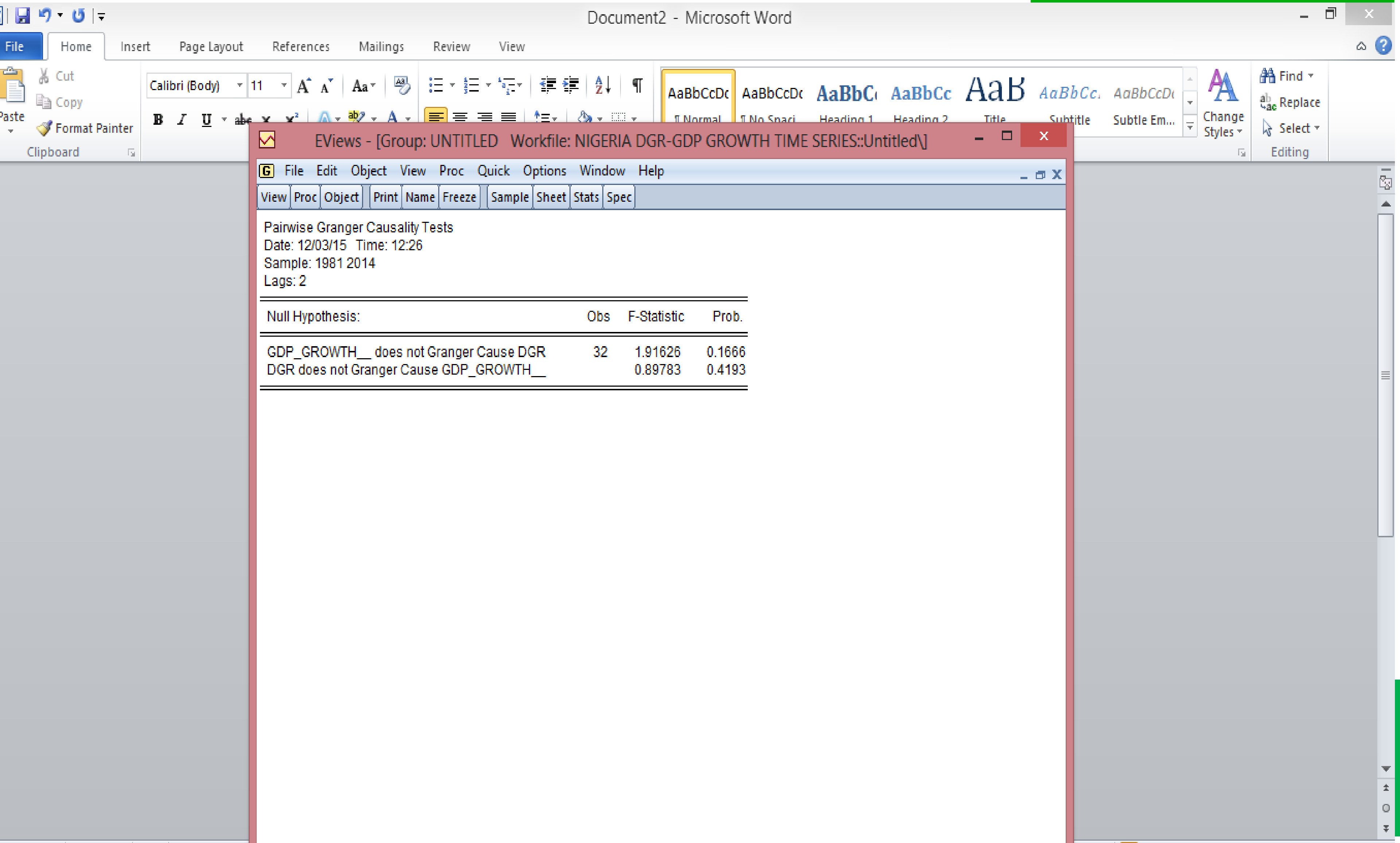
Opening the EViews windows for
Granger causality test operation

BAZE
UNIVERSITY
ABUJA



Granger causality test operation EVIEWS output

BAZE UNIVERSITY ABUJA



Pairwise Granger Causality Tests
Date: 12/03/15 Time: 12:26
Sample: 1981 2014
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
GDP_GROWTH__ does not Granger Cause DGR	32	1.91626	0.1666
DGR does not Granger Cause GDP_GROWTH__		0.89783	0.4193

DGR and GDP Growth in Nigeria: Results of Granger Causality Test

BAZE UNIVERSITY ABUJA

Null Hypotheses	Observations	F-Statistic	Probability	Decision
GDP growth does not Granger-cause DGR	32	1.91626	0.1666	Accept
DGR does not Granger-cause GDP growth	32	0.8978	0.4193	Reject

Summary of findings

BAZE UNIVERSITY ABUJA

i. Both DGR and GDP are stationary based on Augmented Dickey Fuller (ADF) test.

ii. The trace test under Johansen co-integration method indicates one co-integrating equation at 5 percent level of significance.

iii. From the VECM result, it is evident that DGR has significant long-run negative impact on economic growth of Nigerian economy. The long-run negative relationship between DGR and GDP growth tested statistically significant (approx. 10%) by a negative coefficient of DGR.

iv. The Granger causality test results showed the presence of one-directional causality; *causality runs from DGR to GDP with no feedback*. This indicates high level of DGR significantly leads to a slow-down in the economy.

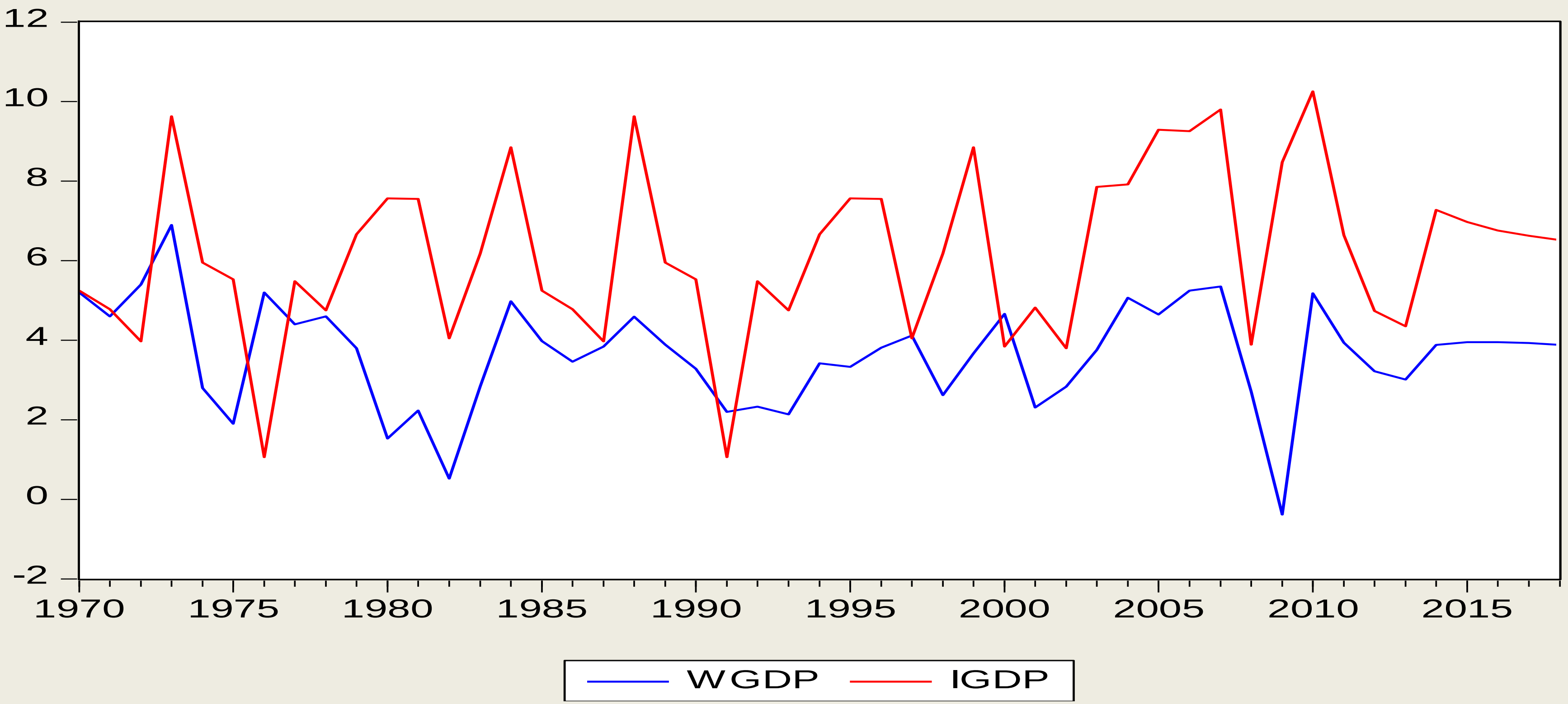
- In essence, the results from the present study align with the theoretical and some empirical positions on impact of sovereign debt on the economy, namely, that output and consumption will grow more slowly than they have, had there been no large sovereign debt (Samuelson & Nordhaus, 2010; Reinhart & Rogoff, 2010; Egbetunde, 2012; Udoka & Ogege, 2012; Reinhart *et al*, 2012).

Insights / implications

BAZE UNIVERSITY ABUJA

- Stop piling up national debts; it may not necessarily grow your economy; rather, diversifying your economy by promoting SMEs across sectors, quality education, rebuilding fiscal buffers, external reserves, introducing investor-friendly policies, could provide better options for achieving macroeconomic stability, sustainable and inclusive growth.

Further illustration…
India & Global growth: any causal nexus?
(Econodata File_6)



Financial Econometrics Know your limits!

BAZE UNIVERSITY ABUJA

- Plurality of models often with similar explanatory ability.
- Analysts' fixation on numbers – non or inadequate discussion of economic / management importance / implications of statistical results
- "Not all problems have solutions" – Microsoft
- It is not a mistake to use logic without statistics; logic does not need empirical verification (Nassim Nicholas Taleb (2004) in his *Fooled by Randomness,* Penguin Books).

Keywords

BAZE UNIVERSITY ABUJA

- ADF, ANNs, ARCH, Autoregression, heteroskedasticity, Bias, Computer Science, Correlation, Econometrics, Descriptive statistics, DF, Financial modeling, Granger causality, EMH, Hypothesis testing, Inferential statistics, Johansen system cointegration test, Macroeconomic model, Mathematics, MSAR, Time series, GARCH, TARCH, Sample period, SETAR, Regression, Technology, Unit root test, VAR, VECM.

To sum up….
Concluding remarks

BAZE UNIVERSITY ABUJA

- ***Financial econometrics*** basically *applies* mathematical, statistical, and computer science techniques to solve problems in finance. It studies how the supply and use of *capital* are considered and measured. ***'****Capital****'*** was defined broadly as any economic resource – a resource or resources that can be used to generate economic wealth; thus, financial econometrics need not be limited to financial capital analytics.
- Time series analysis is an increasingly widely used econometric tool supported by technology to help in gaining faster, quality, and useful insights from observational data on diverse phenomena, not just on finance. Thus, the imperative for protocol /procedure compliance in conducting the requisite statistical trials was stressed.
- Review of a number of empirical cases and problems in finance and macroeconomics with particular emphasis on Granger causality analysis, showed the capacity of technology to make things relatively easy for the discerning business analyst. The need to adopt the parsimony principles in making sense out of the numerous computer outputs was emphasized.

Stephen ARO-GORDON, Ph.D.

Baze University Abuja, Nigeria
Department of Financial Mathematics
Faculty of Computing & Applied Sciences
Email: stephen.aro-gordon@bazeuniversity.edu.ng / getyouthsworking@yahoo.com

BAZE UNIVERSITY ABUJA

December 9-10, 2015 @ BIET, Davangere, India