

**MICROECONOMETRIC MODELING IN APPLIED ECONOMIC RESEARCH:  
THE PAINS, PITFALLS AND PARADOXES**

**INAUGURAL LECTURE**

**PRESENTED BY**

**PROFESSOR ABAYOMI SAMUEL OYEKALE  
{BSc., MSc. PhD (Agricultural Economics), Ibadan}**

**ON**

**14<sup>TH</sup> MARCH 2017**

**AT**

**NORTH-WEST UNIVERSITY MAFIKENG CAMPUS  
SOUTH AFRICA**

## **Introduction**

The Campus Rector, Vice Rector (Teaching, Learning and Quality Assurance), Vice Rector (Research and Planning), Campus Registrar, Other Principal Officers, Deans of Faculties, Directors of School, Members of Campus Senate, Academic and Non-Academic Colleagues, Distinguished Learners, Ladies and Gentlemen.

It is delightful to stand before this distinguished audience to present this inaugural lecture. In this lecture, I will give an account of my stewardship as an academic already promoted to the rank of full professor. As a once in a life time event, I have realized the paucity and pragmatism attached to this lecture, which have also injected in me some sense of academic thoughtfulness for making some meaningful impacts. Although my journey in the lonely and terrific wilderness of research continues, I consider this lecture as a stop-over to give some salient accounts of my academic journeys in the light of my academic achievements. This assignment will help me to re-evaluate my academic pursuits and be reenergized for the greater tasks ahead.

Distinguished guests, though trained as an Agricultural Economist, beyond the theoretical issues I learnt within the four walls of the University, I had practical knowledge of farming from my youthful days. I grew up in an environment where our workplace was the farm with several associated environmental hazards. During those years, I watched my late parents struggling with uncertainties in respect of farm yields. These challenges, coupled with tediousness of farming - especially the cutlasses and hoes type - made me to focus on my studies as a way of avoiding stepping into my father's shoes. In addition to not having access to all the farm inputs that we needed, one of the major challenges we faced was inability to know what quantity of food should be produced in order to cater for our needs all year round. This was also perfectly mitigated by several risks and uncertainties that are associated with farming. My father used to have records of all inputs and output used on the farm, but we often wondered how reliable his data were. For instance, he rarely thought that what we ate out of farm produce was also part of his outputs. This scenario, ladies and gentlemen was the beginning of my exposition to research and its associated problems.

The Campus Rector, when I took "Prof" as my nick name in 1985, I did not even have any strong hope of gaining admission into University. Similarly, I found myself in the profession of Agricultural Economics as an art of God. Back then in my home town – Ipetumodu – I had some friends in my secondary school. We used to have what was then called "Prep" after the normal school work between 3.15 pm and 5.00 pm. There was this day when I returned from "Prep" with three other friends and the question came up from one of us: "Which course should we study?". After deliberating a while, we all agreed to study "Agricultural Economics". By this time, I knew nothing about this course but just based my faith on the agreement reached. It is surprising to note that out of all of us today, I am the only person that read Agricultural Economics. More importantly, too, I am a Professor. This triggers the faith in me that "*To this end was I born, and for this cause came I into the world*" (John 18:37).

Distinguished audience, I have carefully assessed my contributions to research and knowledge in the general field of Agricultural Economics. Majority of my research can be broadly divided into

three areas which are Environmental Economics, Health Economics and Development Economics. Broadly described, my papers on environmental analyses dealt with land use dynamics, sustainable land management, deforestation, vulnerability and impacts of climate change and energy use for sustainable human and economic development. My researches on poverty and inequality focus on factor component and regression-based analytical methods for decomposing sources of income inequality, multidimensional poverty, benefit incidence analysis of government expenditures and pro-poor growth. On analyses of health issues, my researches focus on risky behaviour change, HIV vulnerability, health risk assessment, health care utilization, malnutrition, malaria prevention and treatment behaviour and health insurance uptakes. In many of these studies, the interrelationships between environment, poverty and health had been empirically explored.

Moreover, I have realized that over the past few years, in all the researches that I conducted, a major factor that enhanced topic selection and successful completion was my knowledge of *Econometrics*. I also understand the fact that application of statistical techniques is a major weakness of many students and researchers. In some organized research institute, there could be a unit for statistics and econometrics analysis. The objective of such unit is to oversee every problem that has to do with data analysis. I have never worked in such environment. Given the fact that Agricultural Economics is an Applied Economics, understanding of Econometrics becomes very imperative. This obviously has formed the hub of my research endeavours in the past years and up till today, the role of Econometrics in my day to day research activities cannot be over-emphasized.

Therefore, this evening, The Campus Rector, Distinguished Ladies and Gentlemen, I stand before this great audience to present this inaugural lecture on “*Microeconometric Modeling in Applied Economic Research: The Pains, Pitfalls and Paradoxes*”. In this lecture, I have taken time to discuss some theoretical issues pertaining to application of econometrics for data analysis. Some cognitive emphases had been made on data problems, estimation problems, types and choice of econometric models. The results and findings I obtained from applying many of the discussed econometric models were also presented as my own portfolio of evidences.

### **Agricultural Economics Versus Applied Economics**

Agricultural Economics as a course started sometimes in the 19th century with emphases on application of economic principles to crop and livestock production (Runge, 2006; Hall, undated). The foundation of the discipline was based on theoretical propositions of classical theorists spanning the period between 1700s to early 1800s. Specifically, the trios of Adam Smith, David Ricardo and Thomas Malthus emphasized the role of land as a critical factor of production, although rapidly increasing population pressure aggravates its degradation with serious food productivity implications (Georgescu-Roegen, 1960; Smith, 1776). Therefore, Agricultural Economics was concerned with analysis of optimum allocation of farm resources. The discipline had grown over the decades with significant expansion in scope of empirical applications. Today, pressing issues on management and use of natural resources, rural economic development and international trade are dealt with in theoretical and empirical studies in

Agricultural Economics. Similarly, Agricultural Economics has grown to become a branch of the broad field of economics that is now being accredited for studying in many world's universities.

Two schools of thought had been identified as giving birth to Agricultural Economics (Runge, 2006). The first was the theory of profit maximization that was proposed by the neoclassical economists. Marketing and other farm organizational problems that came to fore during the late 1800s US economic depression formed the second school of thought (Hall, undated). However, significant expansion was witnessed by Agricultural Economics in the 1960s beyond the conventional production theories and farm management and agricultural production to emerging fundamental problems that border on welfare economics and natural resource use management. Consequently, the discipline became more prominent in vital international development policy debates (Runge, 2006; Hall, undated).

Several authors have defined Agricultural Economics in different ways. Specifically, it can be defined as the application of principles of economic to the operations of the agricultural industry (Martin, 1978). It studies how scarce resources are allocated within the farm industry. As an applied social science, it is concerned with the ways in which farm products are produced, distributed, and consumed.

Agricultural Economics research seeks to answer real world questions, and to emphasize testing economic theory against available evidences. While this may limit the contributions of Agricultural Economics to directly extending the bounds of economic theory, in many cases Agricultural Economics research on real world questions has led to vital theoretical contributions. The Agricultural Economics research philosophy, however, tends mostly to result in contributions to the methodology of measuring economic phenomena and testing economic theory (Houck, 1986). Equally important, it results in economic research that is relevant to those outside the Economics profession, to the direct and indirect industry users of economic analysis.

Some departments of Agricultural Economics have recently changed their names to Department of Applied Economics in order to fully reflect what the profession does by testing economic theories and principles on some fundamental behaviours of the agricultural firms and households. Obviously, Agricultural Economics as an Applied Economics deals with practical verification of economic theory and applied econometric modeling in addressing some vital issues within the society. For instance, while economic theory says that the higher the price, the lower the quantity consumed, an Agricultural Economist ventures into verification of this theory by collecting primary data on consumption expenditures and prices and fits the data using econometric models.

## **Identities of An Agricultural Economist**

The profession of Agricultural Economics is of great relevance to national and international policy discourse. The argument that often comes up is whether s/he is an Agriculturist or Economists. Agricultural Economist possesses some basic knowledge of agriculture - Crop Science, Soil Science, Animal Science, pest and disease control, Forestry, Fisheries, Wildlife, Floriculture, Agricultural Extension etc - and integrates these with fundamental knowledge of

economic theories and statistics to address day to day policy issues for promoting agricultural development. Specifically, while a Crop Scientist may be concerned with how to come up with an hybrid variety with certain attributes, an Agricultural Economist thinks about the level of economic gains to be realized if a farmer decides to plant the crop. A Soil Scientist could be interested in the rate of soil degradation, but Agricultural Economist is interested in what are costs associated with land degradation in terms of yield losses and extra cost for procurement of fertilizers.

This does not imply that Agricultural Economics is all about “Economics of this and that”. The discipline has over the years grown in leaps and bounds with substantial multidisciplinary collaborations and integration. Today, the professional knowledge of an Agricultural Economist is sought in virtually all walks of life and the spectrum of research conducted had significantly widened. These include agricultural production and farm management, agri-business management, natural resource management, environmental impact analysis, agricultural finance, food and health economics, international trade, development economics, agricultural policy analysis among others. Similarly, with growing emphasis on promotion of economic development through annexing of the potentials of agricultural activities within and around cities, research in Agricultural Economics had expanded beyond the initial rural focus to urban development issues.

Among the taught courses in Agricultural Economics is Econometrics, which unequivocally forms the focal point of this lecture. My choice of econometrics is not accidental. Having being given the foundation by Professor Saa Ditto in 1993, I had made several efforts to dig into some fundamental knowledge that could assist in my day to day research activities. A quick reference would be some comments at our Departmental Seminar in University of Ibadan where somebody would say “the model you have chosen is not appropriate for your study”. It was then obligatory of the researcher to understand different applications to which econometric models could be put and decide which of them to be used for intended study.

## **Definitions of Econometrics**

Econometrics as a term was first used by Polish Economist, Paweł Ciompa in 1910 (Savoiu and Manea, 2013).. However, Ragnar Frisch who was a Norwegian economist laid the foundation stones of the discipline which has now grown in leaps and bounds over the past few decades (Frisch, 1936). In economic literature, several definitions had been proposed for econometrics. Samuelson *et al* (1954, p. 142) defined it “as the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference”.

Cowles Commission noted that it is “a branch of economics in which economic theory and statistical method are fused in the analysis of numerical and institutional data” [Hood and Koopmans (1953, p. xv)]. Malinvaud (1966) submitted that “every application of mathematics or of statistical methods to the study of economic phenomena” has some econometrics flavours. Christ (1966) hinted that econometrics’ objectives hang around “production of quantitative economic statements that either *explain* the behaviour of variables we have already seen, or

*forecast* (i.e. predict) behaviour that we have not yet seen, or both". Chow (1983) defined econometrics "as the art and science of using statistical methods for the measurement of economic relations".

Given the above definitions, it can be emphasized that econometrics is a discipline that integrates economic theory, mathematical theory and statistics as a perfectly unified discipline. It can be distinguished from any of these disciplines because it applies mathematical and statistical concepts to verify economic theories. It employs some mathematical approaches to measure the magnitudes of parameters that have been postulated in theoretical economics. However, such parameters are sometimes reckoned with only if some statistical tests have confirmed their significance. In some instances, the parameters are objects of further tests in order to ascertain some economic dynamics without merely rushing into some unverified conclusions.

In mathematical economics, some economic theories are postulated mathematically. Econometrics is discussed when some numerical values have been attached to variables in economic theories and such have been subjected to appropriate statistical tests for significance. Therefore, economic theories, mathematics and statistics are necessary conditions for studying econometrics but they are not sufficient.

Econometricians are more or less like perfectionists. They seek to ensure the validity of every underlying assumption. This is to ensure that estimated parameters possess some qualities, which ultimately determine their overall usefulness. Such perfectionist behavior was panoptical of Ragnar Frisch when it was written of him: "his unpublished works are more in number compared to his published ones, mainly due to his perfectionist nature"<sup>a</sup>. In some other instances, due to their perfectionist nature, econometricians were referred to as "lunatics". Shil (1991, p.257) noted that until the 1960s, majority of the "mathematical economists or econometricians were considered part of the lunatics fringe and outside the main stream of economics".

### **Sub-Divisions of Econometrics**

There are two sub-divisions of econometrics. These are theoretical econometrics and applied econometrics. Theoretical econometrics makes use of more of mathematical statistics. It promotes methodologies for testing economic theories through some modeling procedures. Theoretical econometrics spells out the underlying assumptions behind econometric methods. However, Applied Econometrics uses tools of theoretical econometrics to analyze problems in some aspects of economics. For instance, consumption function, investment function and production functions are all derivatives of classical theoretical econometrics, but now streamlined to address some specific economic theories.

Based on the nature of data used, the broad discipline of Econometrics can be divided into two. These are micro-econometrics and macro-econometrics. Micro-econometric analysis deals with "the analysis of individual-level data on the economic behavior of individuals or firms." (Cameron and Trivedi, 2005, p.3). On the other hand, macro-econometrics deals with application of econometric models for analyzing problems that are related to the aggregate economy (Stock, 2001). Such models are run with time series data.

## **Data in Econometric Research – The Pains**

The relevance of econometrics in many social science researches today is motivated by imperfections in available data. Socio-economic data are the concentric of the world which applied economists seek to explain. On the other hand, however, problems emanating from the data are sources of knowledge advancements through formulation of new esoteric models. Therefore, the pains of handling bad data often cannot be compared with associated joy, if ground-breaking solutions are eventually found. Supposed badness of existing data emphasizes the necessity of having comprehensive and representative dataset. Whether they are to be collected by the researchers or they had been collected by someone else, the data must be of sufficient scope and substantially reliable.

Data problem represents one of the foremost obstacles in many social science researches. Previous experiences of enhance ability to handle “second hand data” with care, caution and readiness to overcome every obstacle. Decision to use those existing data is often motivated by several issues. Part of these are the need to respond to some call for papers with very close deadlines, inability to secure the needed funds and time to coordinate and collect data of our own and pressing reality of the “*publish or perish*” syndrome. Most of the times, the divide between data user and data owner is too wide to guarantee any successful communication in the course of using the data. The owner of the data may never have thought that such a study being proposed by the researcher could come from the data. Going through such wilderness of academic rigours may be somehow challenging. The major issue at stake for the researcher is to bring out some handful of useful grains out of the chaffs (Intriligator, 1983).

So many issues often crop up in the process of using such data. In many instances it is even difficult to lay hand on the questionnaire. The coding format may not provide comprehensive information about the variables. Also, while the data might have been collected at household level, researchers may be interested in analysis at individual levels. Some variables such as income which may have been collected at individual level may need to be provided at household level. Missing variables, missing responses to a question and presence of outliers may also constitute barriers. These, and lots of other challenges may crop up and undermine the use to which such the data could be put.

The researchers are to ensure that where possible, every effort is put into collection of comprehensive data. This would involve design of questionnaire aligned with research objectives. Also, the questionnaire must be properly pre-tested by trained enumerators. Inability to do this can lead to omission of important variables and constitutes a lot of hassles on the field. To ensure quality job, it is important for the researcher to personally supervise the data collection processes. A lot of money and time would ordinarily be needed. If a researcher has not secured substantial funding, this is going to be a very difficult task. Even where funding is available, the technical skills required for supervising data collection may not be possessed by the researcher. Essentially, research conduct with “second hand data” may be full of many other challenges. Because several large data sets can be obtained from international organizations like Food and Agriculture Organization (FAO), International Food Policy Research Institute (IFPRI), the World Bank and Demographic and Health Survey (DHS), the scope of available data is often

very wide. Therefore, a researcher may venture into conducting studies of a very large regional coverage. Sometimes, beyond the full grasp of necessary econometric models, differences in climatic conditions, topography, culture, economic challenges and policy processes across those countries within the selected region may constitute some limitations for synchronization and generalization of findings.

In some other instances, due to our busy schedules, the practice of giving questionnaires to a third party who is going to assist with administration can result in cheap and defective research. In some instances, those enumerators could sit down in their living rooms and fill in “imagined responses” for respondents who may be several kilometers away. When such data are subjected to econometric analysis, there are always a lot of problems. More often, there is no remedy thereby taking us back to square one. As data analyst, I had faced situations where data could not produce any meaningful results. Some of those data may be dubious since they were guesses from an enumerator who never went face to face with respondents. This goes against the tenets of academics integrity.

Over the past few years of my academic pursuits, I have got to think first and foremost about data. I had gone through a lot of painful experiences. Sometimes, the problem being investigated would have to be aligned with available data. Of course, it is of no use proposing a study for which existing data cannot perfectly fit. Therefore, the job of the researcher is to be familiar with existing data sets, their scope and sample size. Once the required permission to use the data is secured, every other ethical issue attached to data use must be fully adhered to.

## **Economics and Econometric Models**

One of the major tasks of applied economists is to build models, which should be reasonable and manageable (Intriligator, 1983). Generally speaking, a model represents perfect simplification of a phenomenon such as a system of operation. The suitability of a model is best evaluated from its ability to explain, predict and control a phenomenon. In furtherance to model building, attempts are often made to evaluate the possibility of expanding certain models such that they produce better understanding of their applications and workability. Economic models succinctly states relationships among economic variables. However, econometric models are often specified in algebraic form. They possess stochastic characteristics since the variables exhibit some random properties. In econometric models, random variable is additively included to cater for problems that are related to omission of important variables, specification errors and measurement errors.

A typical multiple regression model can be stated as follows:

$$Y_i = \alpha + \beta_k \sum_{k=1}^m X_{ki} + u_i \quad 1.$$

Where individual observation is denoted as  $i$  and  $k$  represents the number of parameters.

The dependent variable ( $Y_i$ ) is endogenously determined and  $X_{ki}$  are the explanatory variables.  $\alpha$  and  $\beta_k$  are the parameters for estimation and  $\varepsilon_i$  is the stochastic error term.

Specifically, the three components of an econometric model are variables, parameters and error term.

**Variable:** This is factor that is subject to change. In social sciences, we use variables to determine if changes in one factor are leading to changes in some other factors. In econometric, we often refer to some set of variables. For instance, the dependent variable is the variable that is being measured in an experiment. But ideally, econometrics assumes that dependent variable is influenced by what is called independent variables. These are typically the variables that represent the value being manipulated or changed. There are several issues involved in determining which variable is dependent and which ones are independent. In basic pure science, this is obviously not a problem because there should have been some experimental setups that produced the final results. However, in social science, it is not that easy. In many instances, the statistical organization collect data on several aspects of the households' economic behavior, which a researcher needs to make use of for some meaningful studies. In certain instances, econometricians consider the relationships as multidirectional, thereby requiring simultaneous equation estimation.

**Parameters:** This term seems to have originated in mathematics and has a number of specific meanings in different fields. In econometrics, parameters are the coefficients of variables that determine their numerical strengths. In equation 1 above,  $\beta_k$  is the parameter. The magnitude or numerical strength of the parameter has a lot of economic interpretations. In a consumption study, they may represent the marginal propensity to consume if attached to income variable.

**Error Term:** We shall discuss a lot on error term in our subsequent lectures. However, error comes into econometric modeling due to the stochastic nature of variables and inability to exhaust the list of all useful variables. In our example in equation 1, what it means is that it is absolutely impossible to have such a perfect model, where for every respondent, consumption had been perfectly predicted by income. For instance, non-inclusion of other variables like tribe, location, education, household size, household composition etc may have constituted a kind of specification error, the sum total of which will be captured as error term.

## Goals of Econometrics

There are basically three pressing goals of econometrics. These are structural analysis, forecasting and policy evaluation (Intriligator, 1983).

**Structural analysis:** The fundamental goal of econometrics is to present the structural models and estimate the values of their associated parameters. In this manner, econometric models are used to quantitatively measure underlying relationships among variables within the system being analyzed. Some structural parameters that could be estimated in econometric models are marginal products, returns to scale, price and income elasticity of demand, marginal propensity to consume etc. (Reiss and Wolak, 2007).

*Forecasting:* Forecasting implies predicting the future values of a variable within the model. Such prediction is made with parameters already estimated within the time period of available data. Depending on the level of reliability of estimated parameters, it is possible to have perfect knowledge of future values of some variables. This is an important goal for econometric analyses.

*Policy Evaluation:* Econometrics assists in the evaluation of policy alternatives by examining the parameters of policy related variables included in the model. There are three alternatives for selecting policy alternatives. These are “instruments-targets approach, the social-welfare-function approach and simulation approach” (Intriligator, 1983).

### **Peculiar Challenges for African Researchers**

#### **Knowledge Gaps**

A very wide knowledge gap exists between researchers in Africa and those in some developed countries. This is further complicated by lack of interest and inability of many researchers to cope with the subject as a result of weak foundation in Statistics and Mathematics. Similarly, in many developed countries, some Econometricians have sound knowledge of computer programming and software development. Therefore, it is very easy for them to implement some econometric procedures by the means of computer programming. Majority of African researchers are unable to do this. Therefore, at best, many of the African researchers in the field of Econometrics are knowledge users and not originators.

#### **Absence of Clear Niche for Econometrics**

Many African universities and institutions have not been able to properly define definite niche areas for Econometrics, either as a department or as principal focus of research. In many of our institutions, Econometrics is a course that is taught within the programmes of Agricultural Economics and Economics, among others. However, in some institutions abroad, Econometrics if offered as a separate department with Bachelor, Honours, Master and Doctoral degrees awarded. The depth of econometric modeling in such institutions is quite deeper than what obtains in many African universities.

#### **Weak Multidisciplinary Collaborations**

Weak collaboration or its complete absence is a major limitation to application of some econometric models by some African researchers. Beyond the rhythms of ordinary regression analysis, several other models exist which could be easily applied to analyses of some cogent issues in several disciplines. Although it is widely acknowledged now that application of econometrics has gone beyond the discipline of economics, its applicability for research in other disciplines is still limited as a result of weak collaboration.

#### **Policy Failure as a Result of Weak Forecasting Ability of Some Econometric Models**

Econometricians are often expected to function as soothsayers within the context of some pressing economic development problems. Policy makers have got to take interventions being suggested by econometricians very seriously. However, inadequate knowledge of econometric modeling often reduces the forecasting ability of econometric models. This can cause a lot of stress.

to the growth processes of an economy, thereby reducing the level of confidence of policy makers.

## Applied Econometric Models

An econometrician can be likened to a physician, who has the obligation of properly diagnosing the ailments being suffered by a patient before administering treatments. The job begins by first understanding the nature of the data, which may be cross-sectional, time series or panel. The routines of econometric modeling are perfectly decoded from the sentiments of data handling. By sentiments, I refer to those “dos and don’ts” that often form the commandments which are often to be observed religiously if unpardonable sins are to be tactically avoided.

Though seriously criticized, Kennedy (2002) highlighted what was tagged the “*Ten Commandments*” of applied econometric analyses. These commandments warned against jettisoning common sense and economic theories, not asking the right and necessary questions, being ignorant of the context of the study being conducted, failing to keenly inspect the data to discover possibility of outliers, being cajoled by complexity of selected models where simpler ones may perform better, not ensuring parsimony in the selected models, not taking quality time to examine the results of data analyses, failing to understand the cost of data cleaning, not being inflexible in the selection of econometric models, being carried away by statistical significance without duly considering practical significance of the results and not expecting to be criticized.

Depending on the nature of available data and the objectives being pursued, the following are among the econometric models I had applied in my research:

### Ordinary Least Square (OLS) Regression

The OLS regression model is among the most popular econometric models for economic analyses. This model provides the premise for explaining the explanatory ability of some variables on the dependent variable. Example of such models is presented below:

$$Y_i = \alpha + \beta_k \sum_{k=1}^m \beta_k X_{ki} + u_i \quad 2.$$

Regression analysis aims at depicting causality not just mere correlation (McFadden, 1999). In many social science researches, the issue of endogenous or exogenous often constitutes a problem. Jarvis and Media (undated) highlighted the steps that are involved in the definition of the variables selected for social science research. These are: clear definition of the research question, identification of independent variables and identification of the dependent variables. It was noted that independent variables must not be related to one another and they can include age, race and gender. However, it was emphasized that variables such as education and income had been used as both dependent and independent variables depending on the phenomenon being studied.

McFadden (1999) emphasized that some educated guesses must be made by econometricians about the way in which the data being used were generated. Non-realistic assumptions about data

generation processes would lead to misleading results. It was noted that large data sets are desirable because they often possess less "statistical noise". Similarly, the job of an econometrician is to ensure that estimated models are properly tested for statistical plausibility. Also, efforts must be made to always use robust statistical approach which would not produce parameters that are inconsistent and biased.

### ***Basic Assumptions of OLS Regression***

Proper understanding of the assumptions underlying OLS regression underscores less abuse of the statistical procedures. The following assumptions should be borne in mind (Hansen, 2015, :

#### *Assumptions about the error term*

- i. The stochastic error term is a random variable that takes either positive, negative or zero values.
- ii. The expected value (mean) of error term in any particular period is zero -  $E(u_i) = 0$ .
- iii. The error terms has constant variance across each value of the explanatory variable. This is known as the homoscedasticity assumption.
- iv. Errors are normally distributed by their mean values -  $u_i \sim N(0, \sigma_u^2)$ .
- v. Errors at different levels of  $X_i$  are independent. The covariance between  $u_i$  and  $u_j$  is zero. Also,  $Cov(u_i u_j) = 0$  for  $i \neq j$ . This is the non-serial assumption or assumption of no autocorrelation.
- vi. The expected value of  $u_i X_i = 0$ . This implies that the error term is not in any way related to the independent variables.

#### *Assumptions about Dependent Variables ( $Y_i$ )*

- i. Dependent variables are random values which are normally distributed and statistically independent.
- ii. The behavior of the dependent variable is very similar to that of error term. This is due to the fact that given  $Y_i = \alpha + \beta X_i + u_i$ ,  $\alpha + \beta X_i$  is approximately constant. Therefore,  $E(Y_i) = E(\alpha) + E(\beta X_i) + E(u_i)$ . This implies that the expected value (mean) of the dependent variable value at a given level of independent variable  $X_i$  is what can be obtained by the regression equation.
- iii. The variance of  $Y_i$  is also constant because  $\alpha + \beta X_i$  is non-random.
- iv. The dependent variable may or may not be measured with measurement errors.

#### *Assumptions about Independent Variable ( $X_i$ )*

- i. The independent variable is a non-random variable measured with or without error.
- ii.  $E(u_i X_i) = 0$ .
- iii.  $E(X_i X_j) = 0$  for  $i \neq j$ . This implies that  $X_i$ s are truly independent. This is the assumption of non-multicollinearity. This also implies that  $Cov(X_i X_j) = 0$

### **Some Common Pitfalls in Econometric Modeling**

There are a lot of pitfalls to be avoided while carrying out econometric modeling. If this is not done, the expectation that our parameters are Best Linear Unbiased Estimators (BLUE) will be compromised. In many instances, researchers are unable to properly clarify how they have addressed some crucial statistical concerns. In some instances, many researchers that use econometric models have no basic knowledge of their applicability. The era of computing and software development similarly amplifies the rudeness of many researchers to established econometric procedures.

### ***Problem of Extreme Data Values (Outliers)***

Outliers are data points with values being extremely higher or lower than other observations within the dataset (Stevens, 1984; Rasmussen, 1988; Jarrell, 1994). Hawkins (1980) noted that “outlier is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism.” A similar concept to outlier is fringelier, which was introduced by Wainer (1976). This depicts data points with values being close to “three standard deviations from the sample mean” (Wainer, 1976). Outliers could result from data coding error or simply that the outlier does not originally belong to the population being studied (Fox, 2009). In some instances, such outrageous values are masterminds of insincere respondents that may deliberately fill dubious figures in the course of questionnaire administration (Dixon, 1950, p. 488; Wainer, 1976).

However, such outliers pose serious challenge to econometric analysis because of the likelihood of distorting the magnitude of estimated parameters (Zimmerman, 1994, 1995, 1998). This is obviously possible due to very high likelihood of increasing the variance of error thereby increasing the chance of statistical significance validity and concomitant commitment of type II error. In addition, depending on the nature of existing data, outliers promote data estimation difficulties by distorting the distribution of the data thereby producing some form of skewness. Also, estimated parameters can suffer substantial biasness thereby leading to some misleading conclusions (Rasmussen, 1988; Schwager and Margolin, 1982; Zimmerman, 1994).

Environmental Protection Agency (2006, p 116) highlighted five steps that should be taken when treating data points suspected to be an outlier. These are:

1. Identification of extreme data points that could be an outlier;
2. Conduct of some statistical tests;
3. Scientific review of outliers and making concrete decision on their statistical disposition;
4. Analysis of the data with and/or without the outliers in order to judge their implications on computed parameters, and
5. Documentation of the entire processes for reference and academic transparency.

### ***Detection of Outliers in Research Data***

There are a number of ways to detect presence of outliers in research data. Detection is important in order to decide on the best way of correcting any associated problem. Iglewicz and Hoaglin (1993) suggested that outliers should be properly identified, labeled for further investigation and

accommodated by using the most appropriate statistical techniques. The methods that can be used for detecting outliers can be divided into the following:

### *Standard Deviation Method*

Standard deviation is one of the classical statistical methods for detecting an outlier. Therefore, by computing the mean and standard deviation of a univariate data, it is possible to detect those points that could be classified as outliers. The analysis is based on the definition of points within which our data points are expected to belong. We can talk about the following points:

1 standard deviation method:  $\mu \pm \sigma$

2 standard deviation method:  $\mu \pm 2\sigma$

3 standard deviation method:  $\mu \pm 3\sigma$

Observation outside these data points can be considered as outliers. Chebyshev's inequality is directly applied, which is based on specification of a random variable X with a mean ( $\mu$ ) and standard deviation ( $\sigma$ ). For any expression where  $k > 0$ , it can be shown that

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad 3.$$

$$(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}, \quad k > 0 \quad 4.$$

If the variable tends to be a normal distribution, its probability density can be expressed as:

$$f(x) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad 5.$$

$$f(x) > 0 \quad 6.$$

and

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad 7.$$

The degree of variance with expected distribution is often judged from the fact that 68%, 95% and 99.7% of the distribution should lie within the first three standard deviation from the mean.

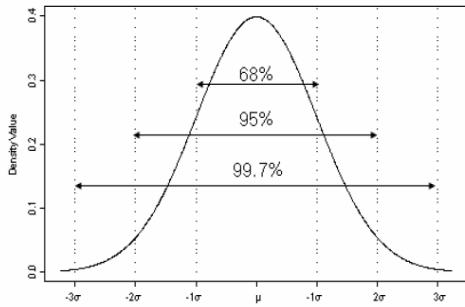


Figure 1: Normal distribution of univariate data and probability distribution based on standard deviation  
Source: Seo (2002).

*Draw the scatter diagram*

By drawing the scatter diagram of two variables within a data set, it is possible to detect those data points with extreme values (extremely low or high). An example of scatter diagram is presented below. The figure shows an outlier datum point marked in a circle.

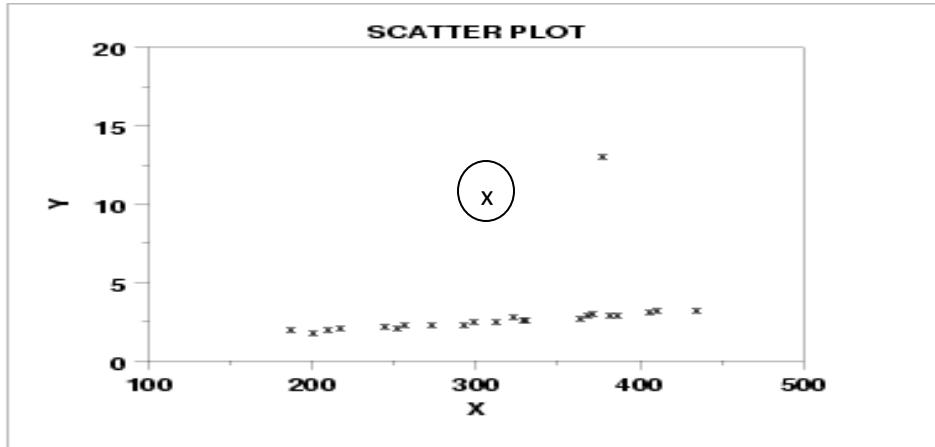


Figure 2: Example of a scatter diagram showing outlier

*Compute Some Descriptive Statistics*

The presence of outliers in a dataset can be detected by computing some descriptive statistics. A cursory look at those statistics can raise some concerns. For instance, a very high standard deviation or variance of continuous data can sensitize analysts to further queries.

*Z-Score Computation*

Outliers can also be discovered using the Z-Scores. All the data points are to be converted to standard scores using the mean ( $\mu$ ) and standard deviation ( $\sigma$ ). The z-score is expressed as:

$$z = \frac{(x_i - \mu)}{\sigma} \quad 8.$$

Where

$$x_i \sim N(\mu, \sigma^2) \quad 9.$$

If the sample size is 80 or less, a data point would be considered as outlier when the absolute value of the computed standard score is  $\geq 2.5$ . Also, if the sample size is greater than 80, a data point is said to be an outlier if the absolute value of its standard score is  $\geq 3$ . Schiffler (1988) had proved that possible maximum values of standard z score is directly related to the sample size ( $n$ ), and it can be expressed as  $(n - 1)/\sqrt{n}$ .

There are some inherent problems with this method. These include likelihood of getting inflated standard deviation value due to presence of some outliers. This leads to masking problem which arises as a result of inability to properly identify an outlier.

### *Use of Modified Z-Score*

In order to avoid masking problem that could result when using z-score due to possibility of inflated standard deviation due to presence of some extreme values, the modified z-score had been proposed in literature. The modified form of the z score is expressed as:

$$M_i = 0.675 \frac{(x_i - \mu)}{MAD} \quad 10.$$

where  $E(MAD) = 0.675\sigma$  for large sample that is normally distributed. The MAD can be computed as follows:

$$MAD = Median|x_i - \mu| \quad 11.$$

Iglewicz and Hoaglin (1993) submitted that data points with modified absolute Z-scores  $> 3.5$  should be labeled as potential outliers.

### *Tukey's Method*

An approach for detecting outliers was proposed by Tukey's (1977). A data point is an outlier if:

$$Y_i < Q_1 - 1.5IQR \text{ or } Y_i > Q_3 + 1.5IQR \quad 12.$$

Where  $Q_1$  = lower quartile,  $Q_3$  = upper quartile, and  $IQR = (Q_3 - Q_1)$  is the inter-quartile range.

### ***Statistical Tests for Detecting Outliers***

Some statistical methods had been proposed for identifying outliers within a data set. The summary of those methods is in table 1.

Table 1: Summary of statistical methods for detecting outliers

Sample size	Test	Normality assumption	Multiple outliers
$n \leq 25$	Dixon's test for extreme value	Yes	No/Yes
$n \leq 50$	Discordance test	Yes	No
$n \geq 25$	Rosner's Test	Yes	Yes
$n \geq 50$	Walsh's Test	No	Yes

Source: Adapted from United Nations Environmental Protection Agency (2006, p 117)

### *Dixon's Test for Extreme Value*

When the samples size is  $\leq 25$ , the extreme value test originally proposed by Dixon can be used to determine whether a data point is a statistical outliers. The test assumes that without the outliers, the distribution of the data would follow normal distribution, and two cases (extremely low and high values) are always taken into consideration. It is then expected that test of normality is carried out with the exclusion of suspected outliers. In case the result confirms non-normal distribution, other more appropriate tests should be used or the data can be transformed into normal distribution.

### *Discordance Test*

This test is able to determine if an extreme value is an outlier or not. This test also assumes normal distribution of the data when the suspected outlier is excluded. It is therefore necessary to check whether normality assumption is violated or not before proceeding with the test.

### *Rosner's Test*

When the sample size is  $\geq 25$ , Rosner's test can be applied for testing and detecting up to 10 outliers. This parametric test assumes that with the exclusion of the outliers, the data will be distributed normally. It is also essential to carry out necessary test to confirm the normality assumption before going ahead with the test. In case normality assumption is rejected, data transformation is necessary.

### *Walsh's Test*

Walsh developed a nonparametric approach for detecting many outliers within a dataset. The test is carried out with large sample size of more than 220 at 5% level of statistical significance. When sample size is more than 60, 10% level of significance is to be used. Because the assumption of normal distribution is not binding, this test can be used with skewed dataset.

### ***Collinear Variables - Multicollinearity***

Another major challenge in regression analysis is collinearity among independent variables. When this problem exists, the classical assumption of  $Cov(X_i X_j) = 0$  is completely violated

with some econometric consequences. Two independent variables are said to be collinear if there is very high correlation between them. Therefore, we are dealing with a case where it could be said that:

$$X_1 = \alpha + \beta X_2 \quad 13.$$

where  $\alpha$  and  $\beta$  are constants.

Some examples of variables that conventionally may show multicollinearity are age of household heads and number of children, years of education and income of households' heads, weight and height of an individual etc. There are cases when researchers deliberately include multicollinear variables like age and its square as independent variables. There should be veritable reasons for such action.

Although multicollinearity is a problem that should be properly evaluated before progressing in econometric analysis, many researchers are either ignorant or pretend that everything is alright with their choice of predictors. The whole exercise then becomes that of "garbage in garbage out".

### *Causes of Multicollinearity*

Multicollinearity can be a problem in econometric analysis due to the following reasons:

1. Inability to use dummy variables properly. This could be as a result of not excluding the reference group.
2. Inclusion of variables that had been computed from other variables already included in the equation. For instance, a model with both household size and per capita household income could be problematic because per capita income was calculated by household size.
3. Inadequate specification of econometric models. This could be functional specification error or inability to understand the principal variables to be included.

### *Effects of Multicollinearity*

Multicollinearity can be detected when the following symptoms are found in the course of data analysis:

1. Significant or high magnitude of changes in the value of estimated parameters resulting from an addition or deletion of an independent variable.
2. Changes in the sign of some parameters as some variables are being introduced into the model.
3. Many of the regression parameters would show no statistical significance even if a high coefficient of determination had been computed.
4. When a parameter does not show statistical significance when estimated with other variables in a multiple regression model but statistically significant if singly estimated in a simple linear regression model.

## *Detection of Multicollinearity*

Many diagnostic indicators have been proposed for detecting the presence of multicollinearity. These are discussed below.

### *Variance Inflation Factor*

One of the ways to quickly detect whether multicollinearity is a problem in econometric analysis is by computing the tolerance level. This is computed from the variance inflation factor (VIF) defined as

$$VIF = \frac{1}{(1 - R_j^2)} \quad 14.$$

$R_j^2$  is computed as the coefficient of determination when  $X_j$  is set as the dependent variable, while other independent variables used as predictors.

It should be noted that the sampling variance of the  $j$ th coefficient  $\beta_j$  can be expressed as:

$$V(\hat{\beta}_j) = \frac{1}{(1 - R_j^2)} \frac{\sigma^2}{(n - 1)S_j^2} \quad 15.$$

Fox (1997) noted that variance of error is  $\sigma^2$  and the variance of  $X_j$  is  $S_j^2$  which can be expressed as:

$$S_j^2 = \frac{\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2}{(n - 1)} \quad 16.$$

It can be proved that VIF measures the effect of multicollinearity on the precision of estimated  $\hat{\beta}_j$ . This is the expression of the ratio of  $\hat{\beta}_j$  variance to the expected value of variance if there is no multicollinearity among the dataset. However, the level of VIF that would produce poor parameter has not been properly defined. When the tolerance level of a parameter is too low, multicollinearity is to be suspected. The tolerance is computed as the inverse of variance inflation factor (see table 2 for computed tolerance levels at different levels of  $R_j^2$ ).

Table 2: Hypothetical tolerance levels at different levels of  $R_j^2$

$R_j^2$	$1 - R_j^2$	VIF	Tolerance
0.99	0.01	100	1
0.95	0.05	20	5
0.9	0.1	10	10
0.8	0.2	5	20
0.7	0.3	3.33	30
0.5	0.5	2	50

0.3	0.7	1.43	70
0.2	0.8	1.25	80
0.1	0.9	1.11	90
0.009	0.991	1.01	99.1

### Condition Index

Another way of evaluating variables for multicollinearity is through computation of the Condition Index (CI). This is expressed as:

$$\text{Condition Index (CI)} = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} \quad 17.$$

Furthermore,

$$CI) = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} = \sqrt{\frac{1 + \sqrt{r_{12}^2}}{1 - \sqrt{r_{12}^2}}} \quad 18.$$

Condition indices of 10 and above are taken to imply significant multicollinearity. For instance, if CI is set at 10 in the equation above,  $r_{12}^2 = .9608$  (Fox, 1997).

### Leamer's Method

Leamer suggested a measure for detecting multicollinearity for a selected variable in the model. This involves computation of a statistic known as  $c_j$ . This is expressed as:

$$c_j = \left[ \frac{(\sum_i (X_{ij} - \bar{X}_j)^2)^{-1}}{(X'X)_{jj}^{-1}} \right]^{1/2} \quad 19.$$

By definition, in the above expression, we compute ratio of variances of estimated  $\hat{\beta}_j$  without and with other independent variables, and later find its square root. If no correlation exists between  $X_j$  and other independent variables, the value of  $c_j$  would be equal to 1.

### Farrar–Glauber test

A method proposed by Farrar and Glauber (1967) to test for multicollinearity is dependent on the outcomes of three independent tests. In the first test, the presence of multicollinearity is examined. The second test determines the collinear variables and the third test finds the form of multicollinearity. Given the assumption that the explanatory variable is a multivariate variable that is normally distributed, the following tests were proposed:

*Chi-Square Test for the Presence of Multicollinearity:* The tested null hypothesis is that the independent variables (X's) are orthogonal in distribution. This can be carried out by computing

a statistic from the determinant of  $|X'X|$ . According to Barlett (1937), transformation of  $|X'X|$  is obtained as:

$$\chi^2 = - \left[ n - 1 - \frac{1}{6}(2p + 5) \right] \ln |X'X| \quad 20.$$

This statistic has distribution defined to be Chi-Square. Also,  $v = \frac{1}{2}p(p - 1)$  is the degree of freedom. Number of observations is denoted as  $n$  and  $p$  denotes the number of explanatory variables. This test is known as the Barlett's test of sphericity with the computed statistic compared with tabulated value. If the former is higher, the null hypothesis denoting orthogonality must be rejected.

*F-Test for Determining Collinear Regressors:* Using this method, the null hypothesis that  $R_j^2 = 0$  can be tested.  $R_j^2$  is the multiple correlation coefficient when we regress  $X_j$  against the other explanatory variables. The statistic to be computed is stated as:

$$\omega_j = \left[ \frac{1}{(1 - R_j^2)} - 1 \right] \left[ \frac{(n - p)}{(p - 1)} \right] = \left[ \frac{R_j^2}{(1 - R_j^2)} \right] \left[ \frac{(n - p)}{(p - 1)} \right] \quad 21.$$

The distribution of  $\omega_j$  is in the form of F-distribution with  $(n - p)$  and  $(p - 1)$  degree of freedom. When  $\omega_j$  is higher than the F table value,  $X_j$  is collinear.

*Testing the Pattern of Multicollinearity Using T-test:* In order to determine the form of multicollinearity that exists within a model, partial correlation coefficients between  $X_i$  and  $X_j$  were used by Farrar and Glauber. The tested null hypothesis states that  $r_{ij.12.p} = 0$ . A t-statistic ( $t_v^*$ ) was computed as:

$$t_v^* = \frac{r_{ij.12.p} \sqrt{n - p}}{\sqrt{1 - r_{ij.12.p}^2}} \quad 22.$$

$t_v^*$  follows student's t distribution, and  $v = n - p$  degree of freedom. If  $t_v^* > t$  from table, then we are to accept that variables  $X_i$  and  $X_j$  are responsible for multicollinearity.

### *Correlation Coefficients between the Explanatory Variables*

High correlation coefficient between variables gives some indications that they are collinear. However, determining what level of correlation coefficient to worry about is sometimes controversial. This is due to the fact that researchers often work towards presentation of a model which exhibits zero tolerance to multicollinearity. A rule of thumb is to compare the computed correlation coefficients ( $r_{ij}$ ) with the coefficient of determination ( $R^2$ ). Huang (1970) noted that multicollinearity is harmful if the former is higher than the latter. However, this condition is sufficient in order to detect multicollinearity (Judge *et al*, 1985).

### *Instability of Estimated Parameters*

Find out how stable the parameters are when different samples are used. Suppose there is dramatic changes in the parameters' values, multicollinearity should be suspected.

#### *Consequences of Multicollinearity*

1. As long as perfect multicollinearity is not found in data sets, OLS estimators are still BLUE (Best Linear Unbiased Estimators).
2. The standard errors of the parameters are higher thereby leading to low t-statistics and a higher likelihood of committing type II error.
3. The computed confidence intervals for estimated parameters will be so high.

#### *Remedies for Multicollinearity*

1. Correct any error in data specification such as the use of dummy.
2. Use common sense to determine those variables that could move together in the context of the population being studied.
3. Select those fundamentally essential variables based on well articulated theoretical framework and literature review.
4. If it is possible, think of increasing the sample size. This will address the problem of multicollinearity resulting from micronumerosity.
5. Factor analysis or principal component analysis can be used to generate indices from some highly collinear variables.
6. Forward or backward regression method can be used to select the best model given that redundant variables would be dropped from step to step or most relevant variables would be added across the steps.
7. Use ridge regression instead of ordinary least square method.

#### *Heteroscedasticity*

This problem arises when the variance of error is not constant across each of the observations. Such problem reflects a situation where the data had been drawn with diverse conditional probability distribution and dissimilar conditional variance. Given a linear regression model:

$$Y_t = \alpha + \beta_1 X_{1t} + \dots + \beta_k X_{kt} + u_t \quad 23.$$

Heteroscedasticity assumption is expressed if

$$\text{Var}(u_t) = E(u_t^2) = \sigma_t^2 \text{ for } t = 1, 2, \dots, n \quad 24.$$

Since  $t$  subscript is attached to sigma squared, it implies that the disturbance error term for each of the observations is obtained from a probability distribution that does not possess similar variance.

#### *Consequences of Heteroscedasticity*

The associated problems in respect of non-constant variance of error term for OLS estimation are as follows:

1. The OLS estimator will be unbiased.
2. The OLS estimator will be inefficient implying violation of the BLUE condition.
3. The variances and covariances with OLS estimators are going to be biased and inconsistent.

4. Results of hypothesis testing will be invalid.

#### *Detection of Heteroscedasticity*

The presence of heteroscedasticity can be detected by using any of the following methods:

#### *Graphical Plot the Residuals*

The square of the error term can be computed and plotted against an explanatory variable strongly suspected to be associated with the error variance. In case many explanatory variables are likely to be related with the variance of error, separate graphical plots should be made for each of them. However, an alternative graph with squared residuals plotted against the values of the dependent variable computed from estimated OLS regression equation can be plotted. Such graphical representations can be obtained from some available software. It should also be noted that graphical presentation is not a formal test for being double sure of the presence of heteroscedasticity. It only raises some suspicions for further statistical investigations.

#### *Breusch-Pagan Lagrange Multiplier (LM) Test*

In order to detect the presence of heteroscedasticity, Breusch-Pagan test can be carried out. This involves estimation of the model with OLS regression and then obtain the residuals. After that, the following model will be estimated as auxiliary regression:

$$\hat{u}_t^2 = \alpha + \beta_1 X_{1t} + \dots + \beta_k X_{kt} + v_t \quad 25.$$

Compute  $LM=nR^2$

In equation 25, number of observations is denoted as  $n$  and coefficient of determination is  $R^2$ . If the computed LM-stat is greater than  $\chi^2_{k-1}$  (number of parameters estimated is  $k$ ) the null hypothesis of non-existence of heteroscedasticity should be rejected. The major short coming of this test is that it assumes there is linear relationship between the dependent variable which is error term squared and the explanatory variables. It also assumes that the error term is normally distributed. If these conditions are invalid, the outcome of the test would be invalid.

#### *The Harvey-Godfrey LM test*

This test is similar to Breusch-Pagan test except that it uses an exponential functional form. In this test, the square of the residual is also to be estimated from OLS and regressed with independent variable as auxiliary regression:

$$\ln\hat{u}_t^2 = \alpha + \beta_1 X_{1t} + \dots + \beta_k X_{kt} + v_t \quad 26.$$

Also,  $LM=nR^2$  is to be computed and be compared with table value of Chi-Square distribution. Suppose the table value is less than calculated LM statistic, reject the null hypothesis. This implies that heteroscedasticity is present. Similarly, the result of the test would be subjected to significant error if imposition of exponential functional form is invalid and if the distribution of the errors term is not normal.

### *The Glesjer LM Test*

In this test, compute the square of the residual and estimate auxiliary regression of it using OLS and absolute values of the error term as dependent variable.

$$|u_t| = \alpha + \beta_1 X_{1t} + \dots + \beta_k X_{kt} + v_t \quad 27.$$

Compute  $LM=nR^2$  and compare it with the table value of Chi-Square distribution. If the Chi-Square table value (degree of freedom =  $k-1$ ), the null hypothesis is to be rejected showing that heteroscedasticity is a problem.

### *The Park LM Test*

This test is carried out by regressing the square of the residual that it uses an exponential functional form. In this test, the square of the residual is also to be estimated from OLS and regressed with independent variable as the auxiliary regression:

$$\ln \hat{u}_t^2 = \alpha + \beta_1 \ln X_{1t} + \dots + \beta_k \ln X_{kt} + v_t \quad 28.$$

Compute  $LM=nR^2$  and compare with compared with table value of Chi-Square distribution ( $\chi^2_{k-1}$ ) ( $k$  is total number of parameters). If computed statistic is higher than table value, reject the null hypothesis. This gives some indications of the presence of heteroscedasticity.

### *White's Test*

This is a general test that could be used for detecting heteroscedasticity. The advantages of the test are highlighted as follows:

1. No model should be estimated based on structural form of observed heteroscedasticity.
2. The error term may not be normally distributed.
3. The method could help to verify whether observed heteroscedasticity has distorted the conventional formulas for computing variances/covariances using OLS estimators.

The test is carried by assuming a regression model with the following functional relationship:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + v_t \quad 29.$$

The error variance is assumed to be functionally related as follows:

$$\hat{u}_t^2 = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{2t}^2 + \beta_5 X_{3t}^2 + \beta_6 X_{2t} X_{3t} + v_t \quad 30.$$

Compute  $LM=nR^2$ , with number of observations denoted as  $n$  and  $R^2$  represents the coefficient of determination from supplementary regression. If  $LM\text{-stat} > \chi^2_{p-1}$  critical, the null hypothesis is to be rejected showing that significant heteroskedasticity exists.

However, there is the need to note some point about White Test. First, when one or more dummy variables are included, perfect multicollinearity should be avoided by excluding some repetitive variables. Researchers must beware of micronumerosity problem when there are a lot of explanatory variables are in the original model. Under this condition, some variables must be

excluded from auxiliary regression. The linear and interaction variables may be excluded while squared terms must always be retained in the auxiliary regression.

### ***Corrections of Heteroscedasticity***

In a situation where evidences of the presence of heteroscedasticity are found, efforts should be made to correct them. Inability to make necessary corrections implies that though estimated parameters are unbiased, they are inefficient. Therefore, the rule of thumb is that appropriate methods for correction should be taken if there is the presence of heteroscedasticity. There are two ways this can be done. The first is to estimate our parameters with OLS and then correct the estimated variances and covariances in order to restore consistency and efficiency. The second method is to use some other estimators beside OLS for estimating the parameters.

Many researchers using econometric models often go with the first alternative. However, there is the need to perfectly understand the nature of existing heteroscedasticity in order to avoid estimation of estimators with worse characteristics.

Some of the methods for correcting heteroscedasticity are discussed below:

#### ***Heteroscedasticity Consistent Covariance Matrix (HCCM) Estimation***

Consistent estimate of the variances and covariances can be obtained by a method developed by White (1980). This is known as “heteroscedasticity consistent covariance matrix (HCCM) estimator” (White, 1980).. In many software, HCCM had been integrated for easy computation.

#### ***Generalized Least Squares (GLS) Estimator***

If heteroscedasticity is confirmed in a regression model, estimated parameters can be BLUE if estimation is done with the generalized least squares (GLS) method. This is also known as weighted least squares (WLS) method. With this method, the model is weighted with error term’s standard deviation for each of the observations. Therefore, suppose we have

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \quad \text{for } t = 1, 2, \dots, n \quad 31.$$

The variance of the error is to be computed as  $\sigma_t^2$  and the standard deviation is  $\sigma_t$ . Each of the variables in [31] is to be divided by  $\sigma_t$ . This will give specification in equation 32.

$$Y_t/\sigma_t = \beta_1/\sigma_t + \beta_2(X_{2t}/\sigma_t) + \beta_3(X_{3t}/\sigma_t) + u_t/\sigma_t \quad \text{for } t = 1, 2, \dots, n \quad 32.$$

Similarly, [32] can be expressed as a transformed model:

$$Y_t^* = \beta_1^* + \beta_2 X_{2t}^* + \beta_3 X_{3t}^* + u_t^* \quad \text{where } t = 1, 2, \dots, n \quad 33.$$

In [33], homoscedasticity is ensured because  $Var(\mu_t) = 1$ , and OLS should be applied without inclusion of intercept in order to get estimators that are BLUE.

### *Weighted Least Squares (WLS) Estimator*

The approach of WLS is similar in application to the GLS previously discussed. This is as a result of utilization of weight to multiply each of the terms in the regression equation. If the weight be denoted as  $w_t$ , the transformed equation will be specified as:

$$w_t Y_t = w_t \beta_1 + \beta_2(w_t X_{2t}) + \beta_3(w_t X_{3t}) + w_t u_t \quad \text{for } t = 1, 2, \dots, n \quad 34.$$

It should be noted that in the GLS estimator,  $w_t = 1/\sigma_t$ . One of the major estimation issues with the GLS estimator is that for it to be used, there is the need to know the variance and standard deviation of the true error for every observation. Since there are a lot of estimation difficulties in respect of true error variance computation, their values are unknown and unobservable. Therefore, application of GLS estimator is infeasible.

### *Feasible Generalized Least Squares (FGLS) Estimator*

Since there is the need to know ( $\sigma_t$ ) in the GLS estimator and we do not know what their true values are, application of GLS estimator is often infeasible. However, it is possible to obtain estimates of  $\sigma_t$  for each sample. Therefore, GLS estimator could be applied with such sample  $\sigma_t$ . If this is done, the estimator is known as the “Feasible Generalized Least Squares Estimator, or FGLS estimator”.

Given the following linear regression model,

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \quad \text{for } t = 1, 2, \dots, n \quad 35.$$

$$Var(u_t) = \sigma_t^2 \quad \text{for } t = 1, 2, \dots, n \quad 36.$$

It could be assumed that the error variance is a linear function of  $X_{t2}$  and  $X_{t3}$ . Heteroscedasticity therefore is expressed with the following model:

$$Var(u_t) = \sigma_t^2 = \gamma_1 + \gamma_2 X_{2t} + \gamma_3 X_{3t} \quad \text{for } t = 1, 2, \dots, n \quad 37.$$

Estimators of the FGLS for  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  can be obtained by following some econometric procedures.

### **Autocorrelation**

This problem arises when the error terms in regression analysis are correlated. This leads to violation of the classical OLS regression assumption that  $Cov(u_t u_j) = 0$ . Autocorrelation is more of a time series data problem. Discussion of this problem would not be so elaborate because this lecture focuses more on micro-econometric analysis although in some studies such as Oyekale (2006a), Oyekale and Yusuf (2006) and Oyekale (2007a) were based on time series analysis.

### ***Normality of the Dependent Variable***

Another major assumption is that the dependent variable is normally distributed. This implies that before OLS would be used, it should be confirmed that the distribution of the dependent variable is normal. The implication skewed dependent variable is that if subjected to OLS, heteroscedasticity may become a prominent problem.

### ***Robustness Check***

This emphasizes the need for researchers to ensure that the structural stability of estimated parameters is guaranteed. The major bone of contention is that estimated model may exhibit serious sensitivity to slight removal of some variables or some observations. Robustness check is to be carried out in order to ensure plausibility of the estimated model. In some instances, researchers find themselves in a pitfall because the results are neither necessary nor sufficient for concluding any structural validity (Lu and White, 2014).

### **Poisson Regression**

This model is to be applied when we have dependent variables that are count of some items. We may think of how many time a farmer was sick during the cropping season, number of snake bites reported by a household, number of times household ran out of food during a year etc. The analysis follows Poisson probability distribution which is of the form:

$$f(Y_i) = \frac{\mu^Y e^{-\mu}}{Y!} \quad Y = 0, 1, 2, 3, \dots \dots \quad 38.$$

We can prove that  $E(Y) = \mu$  and  $r(Y) = \mu$  which are fundamental properties to be fulfilled with Poisson probability distribution. In this case, the variance is the same as the mean.

In Poisson regression, we try to estimate a model of the form:

$$\ln Y_m = \alpha + \beta_j \sum_{j=1}^k X_k \quad 39.$$

In equations 39, the parameters to be estimated are  $\beta_j$  and  $\alpha$ . The dependent variable is denoted as  $Y_m$ , while independent variables are  $X_k$ .

There is the need to verify if equation 39 Poisson estimation using deviance goodness of fit statistics. This statistic can also help in evaluating how the model had performed. Deviance goodness of fit statistics can be expressed as:

$$\text{Deviance} = 2 \sum_{i=1}^n y_i \ln \left[ \frac{y_i}{\mu_i} \right] - (y_i - \mu_i) \quad 40.$$

In equation 40, n is the number of observations. If the model shows statistical significance ( $p < 0.05$ ), we must conclude that assumption of Poisson distribution had been violated. Corrections of the model can be effected with inclusion of previously omitted relevant independent variables. If this cannot be done, the researcher is permitted to find some other

alternative models such as negative binomial model. Negative Binomial model assumes over-dispersion is gamma distributed across the means. This is also known as Poisson-gamma model. Also, suppose it is assumed that over-dispersion is normally distributed, Poisson normal model should be used.

Another problem that could be encountered with Poisson regression is truncation in which many of the respondents reported zero value. By providing a kernel density distribution graph, researchers should be able to decide whether the data have excess zeros or not. However, the starting point is to decide, using appropriate statistical procedures, whether or not the model fits Poisson regression properly. There are other forms of Poisson regression which include zero-inflated Poisson model, Negative Binomial model, Truncated Poisson model, Panel Poisson model, Multi-level Poisson model, Generalized Negative Binomial model, Zero-Inflated Negative Binomial model, Truncated Negative Binomial model, Panel Negative Binomial model, and Multi-level Negative Binomial model.

#### *Diagnostic statistics with Poisson Regression*

The following are the diagnostic statistics that could be used to evaluate performance of estimated Poisson model:

- a. LR –Test
- b. AIC and BIC
- c. McFadden Pseudo R<sup>2</sup>.
- d. Predicted probabilities.
- e. G<sup>2</sup> (Sum of deviance).

## **Probit and Logit Models**

These models are widely used in applied economic research. They are used to estimate the covariates of some economic behaviours with data collected in qualitative format. Some of the applications of these models include adoption of a technology, use of certain advisory service, willingness to pay for a service such as health insurance, participation of women in the labour market, farmers' access to high value market, market participation of farmers etc. The farm household or the individual has to give either a yes or no to such issues raised above. That is why Probit and Logit models are alternatively called binary dependent variable models with the dependent variable taking the value of 1 for a yes response and 0 otherwise.

### **Logit Model**

Logit model uses cumulative density function  $F(X'\alpha)$ .

$$F(X'\alpha) = \frac{e^{(X'\alpha)}}{1+e^{(X'\alpha)}} = \frac{\exp(X'\alpha)}{1+\exp(X'\alpha)} \quad 41.$$

The predictive probabilities are limited between 0 and 1

## Probit Model

In Probit model, the cumulative density function of standard normal distribution is denoted as  $F(x'\alpha)$ .

$$F(x'\alpha) = \beta(x'\alpha) = \int_{-\infty}^{x'\alpha} \beta(z) dz \quad 42.$$

The predictive probabilities are limited between 0 and 1

The models are estimated using the maximum likelihood methods. It is possible to specify the marginal parameters as:

$$\delta p/X = \frac{\sum F'(X'\alpha)}{n} \alpha_j \quad 43.$$

## Multinomial Logit Model

A modification to logistic regression was proposed by MacFadden (1974). Probit and Logit models deal with a situation where only two categories are available for the dependent or outcome variable. This analysis can be extended to situations where there are more than two groups. For instance, we could face a situation where households may have to choose between four types of medical care providers – traditional doctor, private health centers, public health centers and self medication at medicine store counters. In this case, we could consider age, household size, income, years of education etc as explanatory variables..

The theoretical propositions for using multinomial regression is similar to that of bivariate dependent variable although there are some new issues. Suppose there is a random variable  $Y_i$ , which could take one of some discrete values 1, 2, 3, .....J. The total number of categories we are having is denoted as J. In the example given above, the dependent variable  $Y_i$  could be coded as follows: 1 for traditional doctor, 2 for private health centers, 3 for public health centers and 4 for medicine store counter self medication. Because the dependent variable is nominal and there is no possibility of any definite ordering, multinomial models are more appropriate.

### *Assumptions*

The model assumes the following:

1. It does not assume that the independent variables are normally distributed, linear in parameters, and have homogenous variance.
2. No respondent has more than one value for the case being analyzed. This means that no one belongs to more than one group.
3. Dependent variable cannot be perfectly predicted from any of the independent variables.
4. There is very low multicollinearity showing that the explanatory variables may not be completely statistically independent.
5. The estimation depends on the independence of irrelevant alternatives (IIA) assumption.

### *Model Specification*

Suppose  $Y_i \in \{0, 1, 2, 3, \dots, K\}$ , 44.

Probabilities of the different categories can be modeled such that:

$$P(Y_i = 1) + P(Y_i = 2) + \dots + \dots + P(Y_i = K) = 1 \quad 45.$$

Using the multinomial Logit model, the probability for different alternatives for  $j = 0, 1, 2, 3, \dots, K$  can be specified as follows:

$$\Pr(Y_i = j) = \frac{\exp(\beta_{j0} + \beta_{j1}X_i)}{\sum_{j=0}^K \exp(\beta_{j0} + \beta_{j1}X_i)} \quad 46.$$

The parameters can be identified up to the point where they are normalized. In order to ensure this, the parameters of one of the alternatives should be set at zero. Therefore, we can have the following expressions for the probabilities:

$$\Pr(Y_i = 0) = \frac{1}{1 + \sum_{k=1}^K \exp(\beta_{k0} + \beta_{k1}X_i)} \quad 47.$$

$$\Pr(Y_i = j) = \frac{\exp(\beta_{j0} + \beta_{j1}X_i)}{1 + \sum_{k=1}^K \exp(\beta_{k0} + \beta_{k1}X_i)}, \text{ for } j \neq 0 \quad 48.$$

We can specify the ratio of the log odd for any two alternatives in relation to the estimated variables as shown in equation 49.

$$\log\left(\frac{\Pr(Y_i = k)}{\Pr(Y_i = 0)}\right) = \beta_{k0} + \beta_{k1}X_i \quad 49.$$

This is the expression for Independence of Irrelevant Alternatives (IIA). IIA is directly related to the assumptions of independence and homoscedasticity. However, this assumption is a convenient condition for model estimation but could be so unattractive in the human behavior point of view.

The parameters of the analysis can be interpreted as the marginal effect that the independent variable is having on the log odds-ratio of another alternative denoted as  $k$  when compared with the baseline alternative denoted as 0. This is made possible by computing the marginal effects of the parameters. These are expressed as:

$$\frac{\delta \Pr(Y_i = k)}{\delta X_i} = \Pr(Y_i = k) \left[ \beta_{k1} - \sum_{j=0}^K \Pr(Y_i = j) \beta_{j1} \right] \quad 50.$$

### **Multinomial Probit Model**

When Probit model is directly generalized, we have Multinomial Probit Model. Usage of this model is a bit different from that of multinomial Logit model in the sense that the strict restriction on fulfillment of IIA assumption as applicable to multinomial Logit model is not enforced. Multinomial Probit model is different from multinomial Logit because the former uses

the standard normal cumulative density function (cdf). The probability of selecting alternative  $j$  by  $i$ th observation can be expressed as:

$$p_{ij} = \Pr(Y_i = j) = \Phi(\beta_{j0} + \beta_{j1}X_i), \text{ for } j \neq 0 \quad 51.$$

It should also be emphasized that when multinomial Probit model is used, the results would be generated in a much longer time. However, the coefficients of the parameters are going to be different by a particular factor from that obtained for multinomial Logit regression, while the parameters of marginal effects are going to be similar.

### **Conditional Logit Model**

This model is used when the data has some choice-specific attributes. A good example is the choice individuals make across different healthcare providers. Most of the time, consumers' decisions are directly linked to the inherent characteristics of the provider such as timeliness of being attended to, expected effectiveness of service rendered, staff attitudes to patient, cleanliness of the hospital environment, cost of the service, etc. Therefore, such analysis should rather use conditional Logit model, instead of multinomial Logit model. In the specification of the model, it is postulated that

$$\Pr(Y_i = j/X_{i1}, X_{i2} \dots \dots X_{ij}) = \Pr\left(\frac{\exp(Z_{ji})}{\sum_{k=1}^K \exp(\beta' Z_{ki})}\right), \quad 52.$$

Where  $Z_{ij}$  is a vector of values for attributes.

It is however possible to have a combination of multinomial Logit model and conditional Logit model. The expression changes to:

$$\Pr(Y_i = j/X_{i1}, X_{i2} \dots \dots X_{ij}) = \Pr\left(\frac{\exp(\beta' Z_{ji})}{\sum_{k=1}^K \exp(\beta' Z_{kj})}\right) \quad 53.$$

$$P_{ij} = \left( \frac{\exp(\lambda' X_i + \theta Z_{ij} X_i)}{\sum_{k=1}^K \exp(\beta' Z_{kj})} \right) \quad 54.$$

### **Ordered Probit Regression**

Ordered Probit regression is used when the dependent variable is ordered. Following Greene (2003), the model can be specified by assuming a function  $Y_i^*$ , which may be observed or unobserved as a variables.

$$Y_i^* = \beta_0 + X_{1i} \beta_1 + X_{2i} \beta_2 \dots X_{ki} \beta_k + \varepsilon_i \quad 55.$$

$Y_i^*$  is denoted as ordinal latent variable,  $\beta_s$  are the parameters to be estimated. Actually, if the latent index is above a certain value, high or low risk level would be reported. The interval decision rule can be expressed as:

$Y_i = 1$	if	$Y_i^* \leq u_1$	56.
$Y_i = 2$	if	$u_1 < Y_i^* \leq u_2$	57.
$Y_i = 3$	if	$u_2 < Y_i^* \leq u_3$	58.
$Y_i = 4$	if	$Y_i^* > u_3$	59.

The probabilities of each of the risk levels can be expressed as:

$$\Pr(Y_i = 1) = 1 - \Phi[X_i\beta - u_1] \quad 60.$$

$$\Pr(Y_i = 2) = \Phi[X_i\beta - u_1] - \Phi[X_i\beta - u_2] \quad 61$$

$$\Pr(Y_i = 3) = \Phi[X_i\beta - u_2] - \Phi[X_i\beta - u_3] \quad 62$$

$$\Pr(Y_i = 4) = \Phi[X_i\beta - u_3] \quad 63$$

The marginal effect can be expressed as:

$$d\Pr(Y_i=5)/dX_i = d\Phi[X_i\beta - u_3]/dX_i = \beta\phi[X_i\beta - u_3] \quad 64.$$

In STATA software, mfx compute, predict(outcome(1, 2, 3 or 4)) command was used to compute the marginal effects with respect to different levels of Y after estimating equation 1.

### Seemingly Unrelated Bivariate Probit Regression

Analysing probabilities of a binary dependent variable requires some econometric decisions and choices are made based on some conventional expectations. Probit or Logit regression are ideal for such modelling. However, the Seemingly Unrelated Bivariate Probit (SUBP) will be best if one of the included explanatory variables is endogenous. Therefore, under this condition, estimated parameters using conventional Probit or Logit regression will be biased and simultaneous equation should be used as discussed by Maddala and Lahiri (2009)

The model can be structurally specified as:

$$N_i = \alpha + \pi M_i + \beta_k \sum_{k=1}^n Z_i + v_i \quad 65.$$

$$M_i = \gamma + \delta_k \sum_{k=1}^n Z_i + s_i \quad 66.$$

$N_i$  and  $M_i$  are latent variables. The dependent variable  $N_i$  should be estimated as dummy variables with values of 1 if yes response and 0 otherwise. Also,  $M_i$  should be estimated as 1 if response is 1 and 0 otherwise. Also,  $\alpha, \pi, \beta, \gamma$  and  $\delta$  are the estimated parameters, while  $Z_i$  are the included covariates. Estimated error terms bivariate normal distribution with  $E(v_i) = E(s_i) = 0$ ,  $\text{var}(v_i) = \text{var}(s_i) = 1$  and  $\rho = \text{cov}(v_i, s_i)$ . In the results from STATA software, Wald statistics which tests statistical significance of  $\rho$  should be used to find out if it is proper to estimate the models jointly (recursively) or not.

### Tobit Regression Model

The classical OLS regression does not apply when there are missing values above or below certain magnitude. This often results as a result of censoring of respondents or truncation of dataset. Suppose A variable, Y has been censored and X had been observed for all observations.. Therefore, for some of the respondents, the value of Y may have been reported as a single value or it is taken to be zero (0).

Suppose  $Y = z$  or  $Y > z$ , Y is left censored. Alternatively, if  $Y = z$  or  $Y < z$ , Y is right censored. Y is said to be truncated when we have observed X only for those non-censored

observations. If OLS is applied, estimated parameters will be inconsistent and biased due to omitted variable problem. The error term would also suffer from heteroscedastic problem.

Tobit model seeks to correct these problems. The model is stated as:

$$Y_i = \alpha + \beta_j \sum_{j=1}^k X_i + u_i \quad 67.$$

With  $Y_i$  censored below or above certain points.

### **Stochastic Frontier Model**

According to Farrel (1957), efficiency analysis is made possible when there is a specification of an efficient production function that is to be compared with specifications of actual performance of a firm. This model that is known as stochastic efficiency frontier model was proposed by Meeusen and Vanden Broek (1977). However, Jondrow *et al* (1982) extended the model with applications to cross-sectional farm production data. Using this method allows estimation of individual farmers' level of efficiency. The model can be expressed as:

$$Y_i = f(X_i, \beta) \exp(v_i - u_i) \quad i = 1, 2, \dots, N \quad 68.$$

Where  $Y_i$  = output of  $i$ th farmer,  $X_i$  = vector of input variables,  $\beta$  = parameter for estimation,  $v_i$  = random error and  $u_i$  is error component that measure the level of technical efficiency. It is possible to compute farmers' level of inefficiency using some software such as FRONT 4.1 which was developed by Coelli (1994).

## **My Contributions: Applications of Microeconometric Models**

### *Ordinary Least Square Method*

Oyekale *et al* (2001) analyzed the factors influencing time allocation by school children to farm and off-farm works in Ibadan using OLS. Four functional forms were specified out of which double log was selected as the best function judging from some econometric, statistical and economic criteria. The results showed that the number of hours devoted to work significantly reduced as parents' income increased ( $p < 0.10$ ). In addition, as the income realized from work by children increased, hours devoted to work also significantly increased. It was *inter alia* concluded that policies to enhance economic status of farm households will reduce the extent of subjecting their children to child labour.

Oyekale *et al* (2003) analyzed the factors influencing profit levels among frozen chicken marketers in Ibadan metropolis. Four functional forms were estimated with OLS regression. The linear functional form gave the best fit and the results showed that as the years of experience increased, profit significantly increased ( $p < 0.01$ ). Also, those marketers that were operating on large scale had profit levels that were significantly higher ( $p < 0.05$ ) than small scale operators by average of ₦133,245.00. The findings emphasized the need to increase the scale of operation of frozen chicken marketers in order to explore higher profitability.

Adepoju *et al* (2013) analyzed small scale poultry farmers' risk coping behavior. The data were analyzed with OLS regression. The results showed that household size ( $p < 0.05$ ) negatively

influenced the risk coping behavior of the poultry farmers while ratio incomes from non-farm activities to total income ( $p < 0.01$ ), amount of capital ( $p < 0.05$ ) had positive influences. The recommended that awareness on benefits of small size family, access to credit facilities, diversification of livelihood activities and improvement in level of education should be promoted.

Oyekale and Onasanya (2003) analyzed the effect of intensification on food crop farmers' revenues in Ogun state, Nigeria. The linear functional form was selected as the best functional form. The results showed increasing the amount spent on seeds and fertilizers by ₦1.00 would significantly increase farm revenues ( $p < 0.05$ ) by ₦5.62 and ₦2.61, respectively. It was concluded that there were possibilities of farmers exploring agricultural intensification through fertilizer and seed usage given persistent degradation of their land resources.

Oyekale (2009) analyzed the impacts of climatic changes on farm incomes and welfare of farming households in Nigeria. The data used were combination of time series weather variables and cross sectional data. Data were analyzed with OLS regression and multicollinearity was tested. The result showed higher variability in rainfall between 1971 and 2003. Results showed that in northern Nigeria, agricultural income was significantly influenced ( $p < 0.10$ ) by rainfall variability, wood gathering time, and water fetching time. Also, in the estimated model for northern Nigeria, variability in rainfall reduced welfare significantly ( $p < 0.01$ ), while that for southern Nigeria had a positive sign. It was concluded adverse consequences of climate change would be minimized if agriculture is taken more seriously, especially among farmers in northern Nigeria.

Oyekale (2012a) analyzed the factors influencing sustainable land use in the Niger Delta region of Nigeria using OLS regression. The data were collected from 428 farm households using stratified random sampling. The results showed that unsustainable land use indices were influenced significantly ( $p < 0.10$ ) by inability to secure fertile land, being affected by climate change, experience of land conflict, market distance, being married and experience in farming. It was concluded that persistent change in climate coupled with land related conflicts in Niger Delta are paramount drivers of unsustainable land use.

Oyekale (2012b) analyzed cocoa farmers' vulnerability to climate change. Factor Analysis was used to compute composite indicators of climate change based on reported impacts of some climatic parameters such as extremely high and low temperature, high rainfall, extremely low rainfall, unexpected delay in commencement of rainfall, unexpected delay in rainfall stopping and stormy rainfall. The covariates of composite vulnerability were analyzed with ordinary least square (OLS) regression. The results showed that years of education, households' dependency ratio, age of household heads, growing cocoa as primary crop, farming as primary occupation, households' member sick during cocoa season, illness resulted into missing cocoa spraying on a due date, number of cocoa farms owned, ownership type of cocoa farms, estimated fraction of cocoa plots that is covered with cocoa trees, cocoa trees' age, cocoa rehabilitation's years, climate impacting households' health, own bicycle, own motor vehicle and access to extension services had statistically significant influence on climatic change vulnerability ( $p < 0.10$ ). It was

recommended that vulnerability of cocoa farmers to climate change would reduce through provision of adequate education on climate change and rural market development.

Oyekale (2012c) analyzed the factors influencing environmental hazard exposure of urban households in Ibadan, Nigeria. Composite indicators of hazard vulnerability were constructed and subjected to OLS regression analysis in order to identify the covariates. From the results, bushy or dirty environment, inadequate urban planning, and poor waste management were critical environmental problems. The regression results showed that being a female household head significantly increased vulnerability to hazards from domestic and air pollution ( $p<0.05$ ), while income significantly ( $p<0.05$ ) decreased households' exposure to domestic hazards and water pollution hazards. It was recommended that in order to ensure environmental safety in Ibadan as a fast growing city, there is the need for serious enforcement of existing environmental laws.

Oyekale (2013a) analyzed food crop farmers' climate change vulnerability in Oyo state. Data were collected randomly from farmers from ten villages and OLS regression was used for data analysis. The tolerance levels of the variables were computed and their high values revealed the absence of multicollinearity. The results further revealed that farmers ability to change their type of crops ( $p < 0.05$ ) and change their source of capital ( $p < 0.01$ ) significantly reduced vulnerability to climate change, while changing crop variety ( $p < 0.05$ ) increased vulnerability. The study concludes that farmers were adapting to climate change despite absence of formal institutions for farm insurance. Improvement in knowledge transfer to better enlighten farmers on better adaptation strategies was recommended.

Longe and Oyekale (2013) assessed vulnerability and adaptation of smallholder cocoa farmers to climate change in Osun State. Data were analyzed with OLS regression. The results revealed that household dependency ratio ( $p < 0.05$ ) had positive influence on farmer's vulnerability to climate change, while possession of television ( $p < 0.01$ ) and contact with extension agents negatively influenced vulnerability. Adaptation to climate change reduced significantly with dependency ratio ( $p < 0.05$ ) and sickness of household members. The study recommended that more of media awareness on climate change should be created and functioning healthcare systems should be provided to address rising health issues arising from unfavorable weather.

Oyekale *et al* (2007) decomposed the sources of income inequality in Nigeria using the OLS regression-based decomposition method that was proposed by Morduch and Sicular (2002). The approach follows the conventional Gini-coefficient inequality decomposition. From the results, the variables that significantly increased income inequality ( $p < 0.10$ ) were residence in urban centers, residence in Southwest Nigeria, household size, formal education of household's head, number of times being ill, salary/wage jobs, engagement in non-farm businesses, access to formal credit and access to informal credit. It was recommended that investment in income generating opportunities that would benefit rural and urban poor would reduce income inequality.

Oyekale (2012d) decomposed used the models proposed by Rao (1969) and Araar (2006a) to decompose the sources of income inequality in Nigeria. The results from the two approaches

were relatively consistent although household size's contributions were higher in Araar's method. Overall, inequality was significantly influenced by residence in urban areas, residence in southwest geo-political zone, engagement in salaried jobs, engagement in non-farm businesses, access to grants and formal credits. It concluded that regulations on women fertility, increased access to formal education and some targeted income transfer programmes have potentials for reducing Nigeria's income inequality.

Ogunsola *et al* (2015) analyzed the factors influencing climate change adaptation among cocoa farmers in Osun state. The indices of adaptation were computed from eleven adaptation options using Principal Component Analysis (PCA). Access to radio, awareness of climate change and household size significantly influenced adaptation ( $p<0.05$ ). The study concluded that promotion of media programmes on climate change will assist in adaptation among cocoa farmers.

It should be noted that as simple as it may look, OLS is the foundation of many other econometric models that are known today. Many of these other models were proposed due to deficiencies in OLS addressing some estimation problems due to the nature of available data and violation of essential assumptions. Some other related studies where similar model had been used include Ogunsola *et al* (2013), Akintayo and Oyekale (2013), Oyekale (2013e) and Oyekale (2014a).

### ***Almost Ideal Demand Systems (AIDS) Model and Simultaneous Equation Model***

Oyekale (2000) estimated the price, cross-price and expenditure elasticities of food demand in Nigeria using the Almost Ideal Demand Systems (AIDS) model. The results showed that the expenditure elasticities was statistically insignificant ( $p>0.10$ ) in all the food classes although with positive sign. Substitution effect dominated in the computed price and cross price elasticities in both the compensated and uncompensated approaches. It was concluded that policy reforms to increase food production, increase households' incomes and reduce commodity prices would go a long way in enhancing households' nutritional status.

Another application of AIDS model is Oyekale (2008a) where demand for different form of land use in Nigeria was analyzed. The results showed that restricted estimation approach better fitted the data than the unrestricted model based on statistical significance of parameters of hecterage variable. The unrestricted model violated the assumption of homogeneity, while both did not conform with symmetry assumption. Estimated parameters indicated that increase in national income increased the total area of land demanded. Also, substitution or price effect significantly influenced land use demand, while complementary relationships was shown by permanent cropland and arable land.

Oyekale (2001) estimated simultaneous equation model for livestock products' demand and supply in Nigeria. The structural equation parameters were computed from the reduced equations based on the Inverse Least Square (ILS) analytical approach. From the results, price was the major factor influencing supply and demand for livestock products. The computed price elasticities were 0.5577 and 0.3309 for supply and demand, respectively. Income elasticity of demand was 1.2862. It was projected that between 1999 and 2005, shortage in supply to the tune

of 17,466 tonnes would be recorded. It was recommended that supply in livestock products could be increased by providing incentives for promoting sustainable livestock production.

Oyekale and Falusi (2002) analyzed the factors influencing deforestation and conversion of land for agricultural uses in Nigeria with Two Stage Least Square (2SLS) method. The results showed that structural parameters of growth rates in tuber yields, permanent cropland, other land, GDP, fuel wood production and Structural Adjustment Programme (SAP) had significant impacts on deforestation ( $p<0.10$ ), while agricultural land expansion was significantly influence by growth rates of tuber yield, other land, GDP, livestock population, human population, fuelwood production and forestland. It was concluded that deforestation and agricultural land expansion can be curtailed with promulgation and enforcement of laws regulation the use of forest resources and research and technological development to enhance resource use productivity.

Ajai and Oyekale (2012) analyzed the interrelationship between access to ITN and malaria morbidity in Nigeria. Data were analyzed with Two Stage Least Square method. The results showed that access to ITN was significantly influenced ( $p<0.10$ ) by presence of a pregnant woman in the household, number of residents, residence in north east, residence in south east, and residence in south-south regions, number of under-5 children, age of household heads, male headship of household, attainment of formal education, among others. Also, probability of reporting malaria morbidity increased with presence of a pregnant woman and number of under-5 children. The study concluded that due to possibility of households to get bitten by mosquitoes outside the ITN, malaria morbidity may not always reduce with ownership and usage of mosquito nets.

### *Applications of Poisson/ Negative Binomial Regression Models*

Oyekale (2013b) analyzed gender role in agriculture, climate change exposure and food security in West Africa's Sahel Belt. The data were analyzed with Poisson and Negative Binomial regression methods. Cash deficiency model was fitted with Poisson regression production deficiency model fits the negative binomial regression. From the results, The results showed that factors that significantly influenced ( $p<0.05$ ) number of months with insufficient cash to buy food included number of household members that were more than 60 years, flooding experience and receipt of assistance due to exposure to adverse climatic condition, perception of more overall rainfall, incidence of drought, perception of low ground water, area of food cropland owned, area of owned grazing land, access to community grazing land, ownership of radio and ownership of television. Also, those factors in the production deficiency model with statistically significant impacts ( $p<0.05$ ) included education, ownership of mobile phone, ownership of television and men doing most of the work in large livestock production, fodder production and wood production. The study concluded that food insecurity would be ensured if there are initiatives for farmers to cope with climate change and fragile nature of land resources in the Sahel belt of West Africa.

Andualem and Oyekale (2012a) analyzed the Addis Ababa Zoo Park's economic values using Travel Cost Method (TCM). Data were analyzed with truncated Poisson model. The results showed that travelling cost, monthly income and the number of households' dependants

significantly influenced visitation to the recreational site ( $p<0.10$ ). It was concluded that initiatives to promote demand recreational services in Ethiopia should consider income brackets of targeted consumers.

Oyekale (2013c) modeled cocoa farmers' perception on climate change and how climate change affected reported sick times in Nigeria. Data were collected from cocoa farmers from Ondo, Osun, Ekiti using multistage random sampling procedure and analyzed using Negative Binomial model. From the results, factors that significantly reduced sick times among cocoa farmers were years of schooling and extremely low temperature ( $p < 0.05$ ), while reduction in cocoa yield and too much rainfall increased it. It was concluded that there is the need to ensure proper education among cocoa farmers on climate change in order to mitigate the impacts of changing weather.

Oyekale (2015a) analyzed sustainable land management in relation to food security among climate change affected farmers. Data were analyzed with Negative Binomial regression. Factors that significantly influence monthly food shortage ( $p < 0.01$ ) included exposure to climatic shocks, growing vegetables, fish production, income from business, formal and informal loans, stopped using a variety, late planting, mulching, (at  $p < 0.05$ ) support from government projects, introduction of new crops, stopped growing a crop in a season, planting flood resistant varieties, use of cover crops, use of improved irrigation and use of integrated crop management. The study concluded that adoption of sustainable land management practices would enhance adaptive capacity of the farmers.

Oyekale (2015b) analyzed the effect of occupation induced stress as a result of climate change on cocoa farmers' sick times in South Western Nigeria. The data were analyzed by Negative Binomial Regression. The results showed that morbidity increased significantly with age of farmers, malaria infection, missing of regular cocoa spraying, reduction in cocoa yields and reduced with cocoa farming experience and attainment of formal education. It was recommended that efforts to assist cocoa farmer in coping with health challenges associated with climate change should include provision of accurate climate forecasts, among others.

Oyekale (2015c) analyzed the determinants of morbidity among households' members that were economically active in Nigeria with the 2013 DHS data. The data were analyzed with Negative Binomial Regression. The results showed that years of formal education, smoking in the house, households' access to electricity, household heads' age, access to clean cooking fuel, having place for hand washing and the number of rooms per person showed statistical significance ( $p<0.10$ ). It was concluded that in order to reduce morbidity as a result of exposure to environmental pollution, more awareness should be created.

### ***Application of Probit and Logit Regression Models***

Oyekale and Idjesa (2009) analyzed the factors explaining adoption of improved maize varieties among farmers in Rivers state, Nigeria. Data were analyzed with Probit regression. The results showed that probability of adopting improve maize varieties significantly increased ( $p<0.01$ ) with farmers' farming experience, years of formal education, use of mono-cropping and use of zero tillage. The study concluded that promotion of formal education would enhance transmission of the impacts of technological innovations among farmers.

Oyekale (2012e) analyzed factors influencing adoption of hybrid seeds in Central Malawi. The data were with Probit regression model. The results showed that probability of adopting improved seeds significantly reduced ( $p<0.10$ ) with attainment of formal education, number of adult females, number of children, land areas owned and market distance but increased significantly ( $p<0.10$ ) with farmers' age, access to pest management information, access to output market information and access to livestock information. It was noted that efforts to enhance food production in Malawi should focus on provision of production and marketing and information among others.

Mabah and Oyekale (2012) analyzed the factors explaining adoption of maize technical packages in western Cameroon. The data were analyzed with Logit regression model. The results showed that adoption of technological packages increased significantly ( $p<0.10$ ) with maize farm size, production for the market and agricultural extension services contacts. The study concluded that promotion of extension contacts would enhance adoption of technological packages in Cameroon, among others

Oyekale (2013d) identified the factors explaining dairy cattle adoption among smallholder farmers in Kenya. Data collected from 251 cattle farmers were analyzed using the Probit regression model. From the results, factors that significantly influenced ownership of dairy cattle include: (at  $p < 0.01$ ); being resident in Busia district (-ve), being married (+ve), number of boys (+ve), number of cattle owned (+ve), (at  $p < 0.05$ ); having food problem (-ve). The study concluded that efforts to reduce persistent hunger and poverty among small holder farmers in rural Kenya would go a long way in enhancing household's decision to keep cattle.

Nwigwe *et al* (2009) analyzed the factors explaining participation of smallholder yam-based farmers in formal marketing in Oyo state. Data were collected with structured questionnaires administered to 227 farmers and analyzed with Probit regression model. Results showed that probability of decision to sell yam in markets increased significantly ( $p<0.05$ ) with farm size, access to information, and membership of cooperative society but reduced with household size and transportation cost.

Oyekale and Oyekale (2010a) analyzed the determinants of chronic and transient poverty in Nigeria by using expected poverty generated from three stage Feasible Generalized Least Square (FGLS) approach. The results of Probit regression showed that although production shocks, marketing shocks, credit shocks and robbery significantly increased probability of being transiently poor ( $p<0.05$ ), they had significant but negative impacts on chronic poverty. In addition, probability of being chronically poor increased significantly with dependency ratio, being retired, agriculture employment and frequency of illness but reduced with urban residence, residence in South East and South West.

Oyekale (2012f) analyzed the factors explaining welfare shocks' exposure in Nigeria. Data were analyzed with Probit regression. The results showed that probability of exposure to welfare shocks decreased significantly ( $p<0.01$ ) with access to improved water, improved toilet, health facility, borehole, agricultural inputs, agricultural produce buyers, consumer goods, being

employed, assets owned and formal credit access. The study recommended marginal reforms in order to reduce exposure to welfare shocks.

Oyekale *et al* (2012a) analyzed the determinants of poverty in rural households of Waterside Local Government Area of Ogun State, Nigeria. Probit regression was used for data analysis. Results showed that farming as primary occupation and household size significantly increased poverty ( $p<0.10$ ), while credit/loan obtained, education and monthly expenditure of household significantly reduced it ( $p<0.10$ ). To alleviate rural poverty, the study concluded that households should have adequate access to affordable and easily accessible credit facilities, among others.

Oyekale and Oyekale (2008) analyzed the factors influencing involvement of farmers in HIV prone behavior in Southern parts of Nigeria. Using Probit model, involvement in risky behavior significantly increased ( $p<0.05$ ) with farmers' age, unbelief in reality of HIV, not showing any concern about HIV, not knowing anything about HIV, not having being warned about HIV, not having assistance on HIV prevention, being confused on HIV and report of HIV incidence in the village. It was recommended that provision of adequate education to farmers on several issues that are related to HIV prevention will assist in reducing involvement in risky behavior, among others.

Other studies where I used Probit/Logit regression models are Oyekale and Oyekale (2009a), Oyekale and Eluwa (2009), Oyekale and Oyekale (2009b), Oyekale and Oyekale (2010b), Oyekale and Ogunsanya (2012); Omonira and Oyekale (2012); Adepoju *et al* (2012a); Oyekale (2012g); Oyekale (2008b), Oyekale (2008c), Oyekale and Obi (2012), Oyekale (2012h), Oyekale and Gedion (2012); Oyekale and Ige (2012); Dare *et al* (2012); Adewole *et al*, (2013); Oyekale (2014a), Taiwo and Oyekale (2013), Mufutau and Oyekale (2013), Oyekale (2014b), Oyekale (2014c), Oyekale (2015d), Oyekale (2015e), Adem and Oyekale (2015), Tejidini and Oyekale (2015), Nedombeloni and Oyekale (2015), Oyekale (2015g) and Oyekale *et al* (2015).

### ***Application of Tobit Regression Models and Heckman Model***

Oyekale and Okunmadewa (2008) analyzed multidimensional poverty in Abia state Nigeria using the 2002-2003 Core Welfare Indicator Questionnaire data. Fuzzy set was used to aggregate different welfare attributes into composite welfare indices which were further analyzed for their covariates using Tobit regression. The results showed residence in rural areas, female headed households and illiteracy significantly increased multidimensional poverty ( $p<0.01$ ), while household size, reduced it.

Oyekale (2012i) analyzed the factors influencing intensity of improved soybean varieties' adoption in Central Malawi. The data were analyzed with Tobit regression. The results showed that access to soil conservation information and endowed land areas significantly increased adoption ( $p<0.10$ ). The study recommended promotion of soil management information, among others.

Oyekale (2013f) used analyzed the factors explaining consumption of livestock products in Nigeria using Heckman model. The results showed that residence in rural areas, size of

households and food expenditure per capita showed statistical significance ( $p<0.01$ ). It was concluded that efforts to promote protein consumption should focus on enhancing incomes of rural households, among others. Other studies where Tobit regression model had been used are Oyekale and Ajesi (2011), Oyekale *et al* (2009), Oyekale and Adeleke (2012), Oyekale and Adepoju (2012), Oyekale and Adesanya (2012), Adepoju *et al* (2012b) and Oyekale (2014d)

### ***Applications of Multinomial Logit and Ordered Probit Models***

Oyekale (2014g) analyzed the tperception of female genital mutilation (FGM) across different tribes and vulnerability to HIV and AIDS risk in Nigeria. Data were analyzed with Ordered Probit regression model HIV risks significantly increased with perceptions that female circumcisions enhances cleanliness, socially acceptable, enhances sexual pleasure and is religiously approved. It was concluded that people should be educated on health risks associated with FGM, among others.

Oyekale (2015f) assessed Malawian mother's malaria knowledge, healthcare preferences and timeliness of seeking fever treatments for children under five. Data collected from the Malaria Indicator Survey (MIS) were analyzed using the multinomial regression models. The results showed that health care preferences were influenced by age, residence in Northern region, mothers education, number of days taken off from work by mother, parents paid for fever treatment, among others. Other studies include Onademeru and Oyekale (2012), Orija and Oyekale (2012) and Andualem and Oyekale (2012b)

### ***Applications of Bivariate Probit and Seemingly Unrelated Bivariate Probit Regression (SUBPR)***

Oyekale (2013g) analyzed fishing folks access to early warning and post flood assistances in Lagos State Nigeria using the seemingly unrelated Bivariate Probit (SUBP) regression. Statistical significance of the likelihood ratio test (rho) implies that the model was appropriately estimated with the SUBP method. Two models representing access to early flooding warnings and being given assistance were considered and the Wald test was significant indicating good fit for the data. The results revealed that factors which significantly influenced ( $p < 0.05$ )access to early warning and post flood assistance were received help during flooding and membership of association. The study concluded that the preparedness and ability of fishing folks to cope with the consequences of flooding were very low.

Oyekale (2014e) analyzed the malaria fever morbidity among pregnant women in Nigeria. Data were analyzed using the Seemingly Unrelated Bivariate Probit (SUBP). The results showed that the choice of women sleeping under mosquito nets was influenced by knowledge of the fact that sleeping under nets could prevent malaria and malaria could be prevented using coils and insecticide sprays while those that influenced fever morbidity were age, currently pregnant, knowledge of mosquito nets preventing malaria, household size and sick of fever. Other studies include Oyekale *et al* (2012a), Oyekale (2012k), Oyekale and Adeyeye (2012), Oyekale (2013h), Oyekale (2013i), Ogunsola *et al* (2014), Oyekale (2014f) and Oyekale (2015g).

### *Application of Stochastic Frontier Model*

Oyekale *et al* (2004) analyzed land management and economic efficiency of food crop farmers in Nigeria. The results showed that the parameters of seed, labour and land areas cultivated were statistically significant ( $p<0.01$ ). Also, inefficiency in food production was significantly reduced by land use intensity, family size, education and use of fertilizers, while its use of organic manure and clean clearing increased inefficiency. Average computed economic efficiency was also low (50.50%). Awoyemi *et al* (2006) analyzed allocative efficiency of male and female cassava based farmers in Southwestern Nigeria. The results showed the presence of allocative inefficiency in cassava production among male and female farmers. Also, allocative inefficiency increased production costs by 10.42% and 23.65% respectively.

Oyekale (2006b) analyzed agricultural intensification and efficiency of food crop farmers in Nigeria's rain forest belt. The results showed that hired labour and land areas cultivated had positive and significant effects on food production efficiency. The inefficiency model showed that land use intensity, farming experience, use of fertilizer, non-exposure to environmental degradation and education reduced inefficiency significantly ( $p<0.05$ ) while crop diversification increased it. Average technical efficiency was 52.92%.

Oyekale (2007a) analyzed agricultural intensification and efficiency of food production in Southwest Nigeria. The results showed that technical efficiency increased significantly ( $p<0.10$ ) with crop diversification, land use intensity, crop rotation and planting of cover crops, while it reduced by use of mulching and application of organic manure. Average technical efficiency was 24.78% and the study recommended significant upward shift in technology in order to explore possibility of increasing farm outputs given that estimated parameter of land did not show statistical significance.

Oyekale (2012a) analyzed the impact of agricultural intensification and sustainable land management practices on food production efficiency in the Niger Delta region of Nigeria. The data were analyzed with maximum likelihood estimates (MLE) efficiency modeling. The results showed that economic inefficiency was influenced by population density, population per hectare forestland, land use intensity, the use of harrowing, mulching of farm plots, use of organic manure, fertilizer application, gender of the farm household heads and marital status ( $p<0.10$ ). Computed average economic efficiency was 75.27%, while 34.81 percent had efficiency levels in the range of 80-100%. Average farm efficiency was highest Akwa Ibom state while Rivers state had the lowest. The study recommended that to sustainable food production in the Niger Delta region would be enhanced by development of research into the use of appropriate land use technologies.

Oyekale (2012i) analyzed effect of access to improved seeds on efficiency of soybean production in Central Malawi. Data collected from 300 farmers with structured questionnaires were used. Maximum Likelihood Estimation (MLE) production frontier model was used for data analysis. Results showed that inefficiency significantly reduced with education and usage of improved maize varieties ( $p<0.05$ ) and no significant technical efficiency existed between local and improved soybean seeds growers ( $p>0.10$ ). It was recommended that research on soybean should

focus on high yielding and early maturing varieties, and promotion of extension contacts that could enhance knowledge of farmers on existing technologies. Other studies with the same methodology include Oyekale (2012l) and Oyekale (2012g).

## **Concluding Paradoxes**

Distinguished ladies and gentlemen, I will end this presentation by highlighting what I consider as paradoxes in econometric modeling. Given the previous reviews, the crux of the matter is that there are several unresolved issues and challenges which econometricians face, many of them look like paradoxical dilemmas. In order to address the contents of this chapter, there is the need to make some allusions to the attributes of good econometric models as previously discussed. The following paradoxes can be identified:

### ***Parsimony Paradox***

Parsimony as an attribute of econometric modeling emphasizes the need for estimated models to be able to explain a lot with a little information. This goes in contrary to the biblical verdict that “to whom much is given, much would be required”(Luke 12:48). However, econometric models are expected to have very high predictive power. In many instances, there are many flaws from onset. Such may include bad data which must conform to our theoretical expectations. In many instances we are conscious of measurement errors, even when we were not responsible for data collection. The case for data collected from farmers could be worse. Some of the farmers are enthusiastic, others are diabolical and having some hidden thoughts of frustrating your mission. Then we pack these “good” and “bad” responses together and econometric models must make some sense out of them. More importantly, our subjects of analysis – human beings - are assumed in economic theory to be rational. Irrational behaviour and attitudes often distort our results. But such could be serious in a situation where the results must be explained in accordance with theoretical expectations. Therefore, it may be true that econometric models would seek to explain what is sometime impossible to explain given the current context and situation. But most of the times, the modeler is a “one man orchestral”, who must dance to the tune of whatever he plays whether good or bad.

### ***Plausibility Paradox:***

Plausibility of model refers to absolute stability of the parameters. This is another critical aspect of modeling. Assessing the stability of our model may not always be that easy. But the fact that we must do it requires concerted efforts. Therefore, in some instances, by seeking such perfection, we commit more and more sins. A modeler testing for structural stability may end up getting more confused than before. Such efforts then neither give a necessary nor sufficient conditions for appraising the performance of the models, It therefore goes to show that in some instances, when an econometrician decides to break the nut of a problem by subjecting it to more heat. He thinks of taking it closer to the sun, only to realize that above some heights, “the higher you go, the cooler it becomes”. Therefore, it sometime happens that “the more we look, the less we see”.

### ***Paucity Paradox***

The fate of econometric model depends so much on its predictive power. Therefore, as a rule of thumb, our models must possess very high explanatory ability. This is very fundamental since most of the times, we are interested in forecasting. The logic of predictive power sometimes becomes over-ruled when one consider some other problems such as multicollinearity. In many instances, we seek to find a proxy for a collinear variable. Actually, such proxy variable may not be among our variables. We may be advised to increase our sample size when it is practically impossible to do so. Therefore, econometrician is confined by paucity of means of getting his problem solved. This is essentially imperative given that the model must have very strong predictive power. The paucity paradox therefore often brings us back the “square one”. The expectations of a researcher become so much distorted by unprecedented situation in which he has found himself. A good example is attempting to model a data and your entire model is completely insignificant. The question then comes, what should be included? Are there other variables?

### ***Prescriptive Paradox***

The modeler is often guided by rules and procedures. These are to be taken like religious rites. But the truth is that we are often overwhelmed by this “holier than thou attitude” at the expense of understanding the relative sensitivity of our models. Economic theories as our guarding foundation open to us the doors of *a priori* expectations. Most of the times, if theoretical expectations have been violated, the model is discarded without even asking ourselves why it is so. The illustrious nature of economic theories often beguiles an econometrician from understanding some structural changes that might be happening within the economy, for which his present models must respond. Often times, therefore, econometrician have been accused of being the only person who could find “a black cat” which never existed inside a black box.

### **Acknowledgements**

To the almighty God belongs every glory. He is my “very present help in times of troubles”. I am grateful because of the power in His words, most especially the inspiration granted to the Psalmist in writing Psalm 121 from where I have derived a lot of strength.

I would like to thank the management at the North-West University, Mafikeng Campus under the leadership of the Campus Rector, Prof. Mashudu Davhana-Maselesele, for giving me several opportunities to express my academic and research potentials. I have also benefitted from inspired counsels and motivations to keep moving whatever the obstacles may be.

So many thanks to my honours supervisor, Late Prof. Mabawonku who was so tough with me in research and my Master supervisor, Prof. SG Nwoko for shedding some lights on research methodology. I owe a lot to my PhD supervisor, Prof AO Falusi, who was also my mentor and often takes me as his son. Thanks for directing my research interest into Environmental Economics. Prof. Saa Ditto and Prof. SG Nwoko were my Econometrics lecturers. Thanks for laying the foundations upon which my interest in Econometrics was built. Thanks to Prof. Femi

Ayadi of the University of Houston-Clear Lake, for inspiring my interest in Health Economics. The advice you gave me in 2010 really materialized.

Several other lecturers of mine at the University of Ibadan such as Prof Akinwumi, Prof. FY Okunmadewa, Prof. A. Adegeye, Prof. Olayemi, Prof. Adekanye, Prof. Adeyeye (NISER), Prof. Akintola, Prof. V. Akinyosoye, Prof. MAY Rahji, Prof. SA Yusuf and Prof. V Okoruwa have contributed a lot to my academic career. Similarly, I enjoyed so much rapport and inspirations from many colleagues and friends such as Prof. OI Oladele, Prof. TT Awoyemi (Papa Awo), Prof. IB OLUwatayo, Prof. U Useh, Prof. E. Ebenso, Prof. Funso Kutu, Prof. Salawu, Prof. Babalola, Mrs Abosede Lawson (Mama Law), Prof. BT Omonona, Prof. O. Oni, Prof. AI Adeoti, Prof. K Adenegan, Dr. Abimbola Adepoju, Dr. Kemi Obayelu, Dr. Adeola Olajide, Mr Law, Pastor Phillips among others. Special thanks to everyone who has graced this occasion. May the almighty God preserve your journeys back to your different destinations.

So many thanks to several bodies that have provided me with research funds and awards. I need to mention the North-West University (South Africa), African Economic Research Consortium (AERC) (Nairobi), Poverty and Economic Policy (PEP) (Canada), International Foundation for Science (IFS) (Sweden), Support Africa International (Canada), Global Development Network (India), Africa Initiatives (Canada), United Nations Development Programme (UNDP), Idachaba Foundation (Nigeria), International Food Policy Research Institute (IFPRI), University of Ibadan and the World Bank.

To my childhood friend, Dr. Kunle Ojemakinde, who is a Pathologist at the Byrd Regional Hospital, USA. Thanks for those words of advice and companionship.

To my family members – Dr. TO Oyekale (my wife) and children (IyinOluwa, Jesutofunmi and Iretimofe) for believing in my career dreams and granting me several unique opportunities to function effectively as a Daddy at home.

To my late parents – Mr Julius Oyekale and Mrs Julianah Oyekale - for believing in my dreams by sending me to University even when it was very difficult financially.

## References

- Adem R.A., Oyekale A.S.(2015) Factors Explaining Smallholder Cattle Farmers' Access to Climate Change Information in Semi-Arid Region of South Africa, *Environmental Economics*
- Adepoju A.O., A.S. Ogunmodede and A.S. Oyekale. (2012a). Household's willingness to pay for malaria medical treatment in Ibarapa East local government area of Oyo State, Nigeria. *Asia Life Sciences Supplement*8: 47-58.
- Adepoju, A.O, A.S. Oyekale and O. Aromolaran (2012b). Rural Households' Demand for Domestic Energy in Odeda Local Government Area (LGA) of Ogun State, Nigeria. *Journal of Human Ecology*, 40(2): 115-123.
- Adepoju, Abimbola O., Omowunmi. A. Timothy and Abayomi S. Oyekale (2013). Risk coping behaviour of small poultry farmer in Ogun state. *Asian Journal of Animal and Veterinary Advances*. 2013, 1-10.
- Adewole, A.T., A.S. Oyekale, A. Adeoluwa and O. Cofie (2013). Farmers' Perception on the Use of Urine for Growing Vegetables in Ibadan. *Journal of Human Ecology*, 41(1):19-23.
- Ajai F.F. and Oyekale A.S. (2012).Households' Access to Insecticide Treated Nets (ITN) and Malaria Morbidity in Rural Nigeria: A Two-Stage Least Square Approach. *Life Science Journal*9(4):3860-3866.
- Akintayo, T.A. and Oyekale, A.S. (2013). Land Tenure Pattern, Farm Investment Decision and Farm Income in Ibadan, Nigeria. *Asia Life Science*, Supplement 9: 355-365.
- Andualem G.M. and Oyekale A.S. (2012). Factors Influencing Households' Preferences for Traditional and Modern Health Care Services in Debre-Birhan, Ethiopia, *Life Science Journal* 9(4):3885-3891.
- Andualem M.G, Oyekale A.S. (2012a). A Truncated Poisson Modeling of Visitors' Use-Values of Addis Ababa Lions Zoological Park, Ethiopia. *Life Science Journal* 9(4):3878-3884.
- Awoyemi, T. T., I. B. Oluwatayo and A. S. Oyekale (2006) Estimating Gender Differentials in Allocative Inefficiency of Cassava Farmers in Southwestern Nigeria: An Application of Stochastic Translog Cost Approach. *Journal of Economics and Rural Development*, 15(2): 73-86.
- Bartlett, M.S. (1937). The Statistical Conception of Mental Factors. *British Journal of Psychology*, 28:97-104.
- Cameron, C.A. and P.K. Trivedi (2005). Microeometrics: Methods and Applications Cambridge University Press, New York May 2005.

- Chow, G.C. 1983. *Econometrics*. New York: McGraw-Hill.
- Christ, C.F. 1966. *Econometric Models and Methods*. New York: John Wiley.
- Dare M, A. S. Oyekale and O.O. Olugbire (2012). Rural Households' Sources and Demand for Cooking Energy During Kerosene Subsidy in Oluyole Local Government Area of Oyo State, Nigeria. *Journal of Human Ecology* 40(3): 255-260.
- Dixon, W. J. (1950). Analysis of extreme values. *Annals of Mathematical Statistics*, 21, 488-506
- Eke B. (undated). *Econometrics: Economic Data and Econometric Modeling*.
- Environmental Protection Agency (2006) Data Quality Assessment: Statistical Methods for Practitioners EPA QA/G-9S. Internet file retrieved on 12 March 2015 from <http://itrcweb.org/FileCabinet/GetFile?fileID=6918>
- Farrar, D. & Glauber, R. R. (1967). Multicollinearity in regression Analysis: The problem revisited. *Review of Economics and Statistics*, 49, 92-107.
- Farrell, M.J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society (A, general)*, 120: 253–281.
- Fox, J. (1997). Applied regression analysis, linear models, and related methods. Thousand Oaks, CA: Sage Publications.
- Fox, J. (2009). Regression Diagnostics. Internet file retrieved on 23<sup>rd</sup> February, 2015 from <http://socserv.socsci.mcmaster.ca/jfox/Courses/Brazil-2009/slides-handout.pdf>.
- Frisch R. (1936), "A Note on the Term 'Econometrics','" *Econometrica*, 4(1), p. 95.Hendry DF. Econometrics – alchemy or science? *Econometrics* 2000; 47: 387-406.
- Georgescu-Roegen, N. 1960. "Economic Theory and Agrarian Economics." *Oxford Economic Papers (New Series)* 12:1-40.
- Greene W.H. (2003). *Econometric Analysis*. Prentice Hall,
- Gujarati, D. N. and D.C. Porter, *Basic Econometrics*, 5th edition, McGraw Hill, 2012
- Hall, S. (Undated). Definitions of Agricultural Economics. Internet file retrieved on 25<sup>th</sup> March 2015 from [http://www.ehow.com/about\\_5421851\\_definitions-agricultural-economics.html](http://www.ehow.com/about_5421851_definitions-agricultural-economics.html).
- Hansen, B.E. (2015). *Econometrics*.
- Hawkins, D.M. (1980). *Identification of outliers*. London: Chapman and Hall.

- Hood, W.C., Koopmans, T.C. (1953). Studies in Econometric Method, Cowles Commission Monograph no. 14. John Wiley, New York.
- Houck, J.P. 1986. Elements of Agricultural Trade Policies. Waveland Press.
- Huang, D.S. (1970). Regression and Econometrics Methods. New York, Wiley.
- Iglewicz B. and Hoaglin D.C. (1993). How to Detect and Handle Outliers. Milwaukee, WI: ASQC Quality Press.
- Iglewicz, B., Hoaglin, D. How to detect and handle outliers. ASQC Quality Press, 1993.
- Intriligator, M.D., 1983. "Economic and econometric models," Handbook of Econometrics, in: Z. Griliches† & M. D. Intriligator (ed.), Handbook of Econometrics, edition 1, volume 1, chapter 3, pages 181-221 Elsevier.
- Jarrell, M.G. (1994). A Comparison of two procedures, the Mabalanobis Distance and the Andrews – Pregibon Statistics, for identifying Multivariate Outliers. Researches in the Schools, 1:49 – 58.
- Jondrow, J, Lovell, C.A.K., Materov, I.S. and Schmidt, P. (1982). On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model. Journal of Econometrics, 23: 269–274.
- Judge, GoG., W.E~ Griffiths~ RoC. Hill, Ho Lutkephohl, and T.C~ Lee (1985), The Theory and Practice of Econometrics~ 2nd edition~ Wiley, New York°
- Kennedy, P. (2002). "Sinning in the Basement: What Are the Rules? The Ten Commandments of Applied Econometrics". Journal of Economic Surveys 16 (4): 569–589. [doi:10.1111/1467-6419.00179](https://doi.org/10.1111/1467-6419.00179)
- Lu, X and H. White (2014). Robustness checks and robustness tests in applied economics. Journal of Econometrics, 178 (1): 194-206.
- Mabah G, Oyekale A.S. (2012). Analysis of Factors Influencing Farm Households' Adoption of Maize Technical Package in Western Cameroon. Life Science Journal 9(4):3841-3845.
- Maddala, G. S. and Lahiri, K. (2009). Introduction to Econometrics. Wiley Publishers.
- Malinvaud, E. 1966. *Statistical Methods of Econometrics*. Amsterdam: North-Holland.
- Martin, J.E. (1978). Agricultural Economics: A Critical Review of the State of the Science. *Southern Journal of Agricultural Economics*, July : 39-42.
- McFadden D. 1999. Robust Methods in Econometrics, Chapter 7, Econ 240B

- McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, ed. P. Zarembka, New York: Academic Press, 105–42
- Meeusen, W. and van den Broeck, J. (1977). Efficiency Estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18: 435–444
- Mufutau A.A. and Oyekale, A.S. (2013). Assessment of Climate Change Vulnerability, Housing Qualities and Health Status of Cocoa Farming Households in Osun State, Nigeria. *Asia Life Sciences*, Supplement 9: 301-313.
- Nedombeloni, R. and A.S. Oyekale (2015) Welfare effect of social grants in Thulamela Local Municipality of Limpopo province, South Africa. *Theory and Practice of Socio-Economic Development Socioeconomica*, 4(7): 1-22.
- Nwigwe, C., Victor Okoruwa, Blaise Nkamleu, Olawale Oni, and Abayomi Oyekale (2009). Socioeconomic Factors Affecting Intensity of Market Participation among Smallholder Yam-Based System Farmers in Oyo North Area of Nigeria. *International Journal of Economic Perspectives* 3(2): 131-140.
- Ogunsola, G.O. And A.S. Oyekale (2013).Determinants of Smallholder Cocoa Farmers' Adaptation to Climate Change in Ile-Oluji/Okeigbo Local Government Area of Ondo State, Nigeria. *Asia Life Science*. Supplement 9: 1-10.
- Ogunsola G.O., Oyekale A. S., Adeniyi O.R. and Popoola O.S. (2014) Indigenous Adaptation To Climate Change Among Cocoa Farmers in Southwest State Nigeria, *Asian Pacific Journal of Environment and Development*. 21/22: 44-60.
- Ogunsola, G.O., O.O. Olugbire, A.S. Oyekale, and F.J. Aremu (2015). Understanding Perception and Adaptation to Climate Change Among Cocoa Farmers in Tropical Condition. *Ethiopian Journal of Environmental Studies and Management*, 8(Supple 1):816-825.
- Omonira O.F. and Oyekale A.S. (2012).Households' Willingness to Pay (WTP) for the National Health Insurance Scheme (NHIS): The Case of Ojo Local Government Area of Lagos State, Nigeria. *Life Science Journal* 9(4):3873-3877.
- Onademeru, S. and A.S. Oyekale (2012). Benefit Incidence Analysis of Public Health Expenditures and Households' Caregiver Preferences in Ogun State, Nigeria. *Life Science Journal* 9(4):123-130.
- Orija, S.O. and A.S. Oyekale. (2012). Stigma, Discrimination and Willingness to Buy Vegetables from People Living With HIV&AIDS (PLWHA) in Rural Nigeria. *Life Science Journal* 9(4):2094-2100.

Oyekale, A. S. (2000): "An Application of Almost Ideal Demand Systems (AIDS) to Food Demand in Nigeria". *Nigerian Agricultural Development Studies*. 1 (2):43-52.

Oyekale, A.S.(2001): *Supply and Demand Equations for Livestock Products in Nigeria: A Simultaneous Equation Approach*. Tropical Journal of Animal Science 4 (2): 117-125.

Oyekale, A.S., T.T. Awoyemi and Y.T. Ajose, (2001). School Children's Participation in Farm and Off Farm Activities: A Case Study of Akinyele Local Government of Oyo State. *African Journal of Educational Research* 7(1&2):101-110.

Oyekale, A.S. and A.O. Falusi (2002). Determinants of Deforestation and Agricultural Land Expansion in Nigeria. *The Nigerian Journal of Economics and Social Studies* 44 (1): 113-128.

Oyekale, A.S., T.T. Awoyemi and O.I. Jaiyebo (2003). Marketing Functions and Determinants of Profits Among Frozen Chicken Marketers in Ibadan– Oyo State. *African Journal of Livestock Extension* .2:19-23.

Oyekale A.S. and T.G. Onasanya (2003).Effect of Smallholders' Land Use Intensity on Food Crop Production in Odogbolu Local Government Area, Ogun State. *The Nigerian Journal of Economics and Social Studies* 45 (1&2):31-40.

Oyekale, A.S., A.I. Adeoti, T.O. Ogunnupe, and T.E. Mafimisebi (2004). Land Management and Economic Efficiency of Food production in Ogun and Delta State, Nigeria. *Bowen Journal of Agriculture* 1(1):88-93.

Oyekale A.S. (2006a). A Dynamic Optimal Control Model for Deforestation and Agricultural Production in Nigeria. *ASSET An International Journal*, 6(1):247-254.

Oyekale, A.S. (2006b). Agricultural Intensification and Efficiency of Food Production in the Rainforest Belt of Nigeria. *European Journal of Social Sciences*. 4 (1): 45-54.

Oyekale A.S. and S.A. Yusuf (2006). An Error Correction Modeling (ECM) Of The Determinants Of Deforestation In Nigeria. *Journal of Sustainable Forestry*. 23:19-36.

Oyekale, A.S. (2007a): Determinants of Agricultural Land Expansion in Nigeria: An Application of Error Correction Modeling (ECM). *Journal of Central European Agriculture*. 8(2007):301-310.

Oyekale, A.S. (2007b). Intensive Land Use and Efficiency of Food Production in Southwestern Nigeria. *Agricultural Journal*. 2(2):174-179.

Oyekale, A.S., A.I. Adeoti and T.O. Oyekale (2007). Regression-Based Approach to Income Inequality Decomposition in Urban and Rural Nigeria. *International Journal of Economic Perspectives*. 1(1):45-54.

- Oyekale, A.S. and T.O. Oyekale (2008). Farmers' involvement in behaviours prone to HIV/AIDS: A case study of southern Nigeria. *Journal of HIV/AIDS and Social Services*. 7 (3):289-303.
- Oyekale, A.S. (2008a) An Almost Ideal Demand System (AIDS) For Forest And Agricultural Land Demand in Nigeria *Pakistan Journal of Social Sciences* 5 (8): 813-819.
- Oyekale, A.S. (2008b). Assessment of farm households' vulnerability to climate change in the Niger delta region of Nigeria. *Tanzania Journal of Forestry and Nature Conservation* 78:5-13.
- Oyekale, A.S. (2008c). Land Degradation, Soil Conservation Practices and Poverty Incidence in Southwestern Nigeria. *Agricultural Journal* 3(6):482-487.
- Oyekale, A.S. and F.Y. Okunmadewa (2008). Fuzzy Set Approach to Multidimensional Poverty Analysis in Abia State, Nigeria, *Research Journal of Agricultural Sciences* 3 (7):490-495.
- Oyekale, A.S. and T.O. Oyekale (2009a) Do Mothers' Educational Levels Matter in Child Malnutrition and Health Outcomes in Gambia and Niger? *The Social Sciences* 4 (1): 118-127.
- Oyekale, A.S. and T.O. Oyekale (2009b). Socio-economic effects of HIV/AIDS and farmers' involvement in risky behaviours in Southern Nigeria. *International Social Science Journal*, 197/198:421-430.
- Oyekale, A.S. (2009a). Climatic Variability and Its Impacts on Agricultural Income and Households' Welfare in Southern and Northern Nigeria Electronic *Journal of Environmental, Agricultural and Food Chemistry*. 8(1): 13-34.
- Oyekale, A.S., M.B. Bolaji and O.W.Olawa (2009). The effect of climatic change on cocoa production and vulnerability assessment in Nigeria. *Agricultural Journal* 4(2): 77-85.
- Oyekale, A.S., and C.G. Eluwa (2009). Utilization of Health-Care and Health Insurance Among Rural Households in Irewole Local Government, Osun State, Nigeria, *International Journal of Tropical Medicine*, 4 (2): 70-75.
- Oyekale, A.S. and E. Idjesa (2009). Adoption of improved maize seeds and production efficiency in Rivers State, Nigeria. *Academic Journal of Plant Science*. 2(1):44-50.
- Oyekale, A.S. and T.O. Oyekale (2010a). An Assessment of Income Shocks and Expected Poverty Dynamics in Nigeria. *The Icfai University Journal of Agricultural Economics*.7 (3):18-50.

- Oyekale, A.S. and T.O. Oyekale (2010b) Application of Health Belief Model for Promoting Behaviour Change among Nigerian Single Youths. *African Journal of Reproductive Health* 14(2): 63-75.
- Oyekale, A.S. and D.O. Ajesi (2011). Households' Exploitation of Non-Wood Forest Products (NWFPs) in Okitipupa Local Government Area of Ondo State, Nigeria. *Journal of Horticulture and Forestry* 3(7):222-225.
- Oyekale, A.S. (2012a). Intensive Land Use, Land Management and Efficiency Of Food Production in the Niger Delta Region of Nigeria. *Electronic Journal of Environmental, Agricultural and Food Chemistry*, 11 (3):1188-204.
- Oyekale A.S. (2012b). Vulnerability of Peasant Cocoa Farmers to Climate Change in South-west Nigeria *Journal of Human Ecology* 40(1): 33-41.
- Oyekale, A.S. (2012c). Factors Influencing Households' Environmental Hazard Exposure in Ibadan Metropolis. *Life Science Journal* 9(4):2520-2527.
- Oyekale A.S. (2012d). Determinants of income inequality in urban and rural Nigeria: A regression-based approach, *Asia Life Sciences Supplement*8: 127-142.
- Oyekale, A.S. (2012e). Factors influencing adoption of hybrid seeds among small holder farmers in Central Malawi, *Asia Life Sciences Supplement*8: 87-96.
- Oyekale A.S. (2012f). Factors Influencing Households' Income Shock Exposure and Coping Options in Nigeria. *Life Science Journal* 9(2):595-601.
- Oyekale, A.S. (2012g). Factors Influencing Households' Willingness to Pay for National Health Insurance Scheme (NHIS) in Osun State, Nigeria. *Ethno Medicine*, 6(3): 167-172.
- Oyekale A.S. (2012h). Factors influencing access to climate forecasts in the Limpopo River Basin of South Africa. *Journal of Food, Agriculture & Environment* 10 (3&4) : 1058 - 1062 .
- Oyekale, A.S. (2012i). Adoption intensity of improved soybean varieties and technical efficiency in Central Malawi. *Asia Life Sciences Supplement* 8: 115-126.
- Oyekale A.S. (2012j). Fuzzy Indicator of Sustainable Land Management and Its Correlates in Osun State, Nigeria. *Journal of Human Ecology*, 39(3): 175-182 (2012).
- Oyekale A.S. (2012k). Assessment of Households' Access to Electricity and Modern Cooking Fuels in Rural and Urban Nigeria: Insights from DHS Data. *Life Science Journal* 9(4):1564-1570.

- Oyekale A.S. (2012l). Climate Change and Cocoa Production Efficiency Losses in Ondo East Local Government, Nigeria. *Life Science Journal* 9(3):726-732.
- Oyekale A.S. (2012m). Impact of Climate Change on Cocoa Agriculture and Technical Efficiency of Cocoa Farmers in South-West Nigeria. *Journal of Human Ecology*, 40(2): 143-148.
- Oyekale A.S. and A.H. Adeleke (2012). Assessment of Land Degradation and Farm-Level Deforestation in The Niger Delta Region of Nigeria *Life Science Journal* 9(4):118-122.
- Oyekale A.S. and Y. A. Adesanya (2012).Climate Change and Urban Children's Health: A Case Study of Ibadan South West Local Government, Nigeria. *Life Science Journal* 9(3):894-899.
- Onademeru, S. and A.S. Oyekale (2012). Benefit Incidence Analysis of Public Health Expenditures and Households' Caregiver Preferences in Ogun State, Nigeria. *Life Science Journal* 9(4):123-130.
- Oyekale, A.S., A.O. Adepoju and A.M. Balogun (2012a). Determinants of Poverty among Riverine Rural Households in Ogun State, Nigeria, *Studies in Tribes and Tribals*, 10(2): 99-105.
- Oyekale A.S., A. M. Dare and O. O. Olugbire (2012b). Assessment of rural households 'cooking energy choice during kerosene subsidy in Nigeria: A case study of Oluyole Local Government Area of Oyo State. *African Journal of Agricultural Research* 7(39):5405-5411.
- Oyekale A.S., A.O. Adepoju. (2012). Determinants of Agricultural Intensification in Southwest Nigeria. *Life Science Journal* 9(3):370-376.
- Oyekale, A.S; Ogunsanya, O.A. (2012). Factors Influencing Households' Access to Portable Water in Rural Nigeria *Life Science Journal* 9(3):2488-2494.
- Oyekale A.S and Kapalasa Eliya Gedion (2012).Rural households' vulnerability to climate-related income shocks and adaption options in Central Malawi. *Journal of Food, Agriculture & Environment* 10 (3&4):1505-1510.
- Oyekale A.S. and R.F. Ige (2012). Urban households' willingness to pay for environmental safety in Ibadan metropolis, Nigeria, *Asia Life Sciences Supplement*8: 69-85.
- Oyekale A.S. and Adeyeye A. (2012).Rural Households' Awareness and Willingness to Pay for National Health Insurance Scheme (NHIS) in Ilesha West Local Government Area, Osun State Nigeria: A Recursive Bivariate Probit Approach. *Life Science Journal* 9(4):2086-2093.

- Oyekale A.S. and S.I. Obi (2012). Farmers' perception of vulnerability to climate change and adaptation strategies in Isuiwuato local government area, Abia State, Nigeria, *Asia Life Sciences Supplement*8: 143-152.
- Oyekale, A.S. (2013a). Analysis of climate change vulnerability among food crop farmers in Lagelu local government area of Oyo state. *Journal of Food, Agriculture and Environment* 11(1):887-891.
- Oyekale, A.S. (2013b). Gender Role in Agriculture, Climate Change and Food Security in the Sahel Belt of West Africa: Application of Poisson and Negative Binomial Regression. *Gender and Behaviour*, 2013, 11(2), 5499-5511
- Oyekale, A.S. (2013c). A Negative Binomial Modeling of Cocoa Farmers' Perception on Climate Change and Reported Sick Times in South Western Nigeria. *Journal of Food, Agriculture and Environment*.11 (3&4): 1975-1981.
- Oyekale, A.S. (2013d). Bivariate Probit Modeling of Factors Influencing FFS membership in East Africa. *Asian Journal of Animal and Veterinary Advances*. 8(7):904-913.
- Oyekale, A.S. (2013e). Analysis of Factors Promoting HIV&AIDS Vulnerability among Married Couples in Nigeria: A Bootstrapped Quantile Regression Approach. *Gender and Behaviour*, 11(2), 5752-5761.
- Oyekale, A.S. (2013f). Heckman Modeling of Factors Explaining Livestock Products' Expenditures among Households in Nigeria. *Journal of Animal and Veterinary Advances*. 12(16): 1325-1331.
- Oyekale, A.S. (2013g). Fishing Folks' Access to Early Warning and Post Flood Assistances in Lagos State, Nigeria: A Seemingly Unrelated Bivariate Probit (SUBP) Modeling. *Journal of Animal and Veterinary Advances*. 12(5): 607-611.
- Oyekale, A.S. (2013h). Factors Explaining Dairy Cattle's Adoption Behaviour among Smallholder Farmers in Rural Kenya. *Asian Journal of Animal and Veterinary Advances*. 8(7):893-903.
- Oyekale, A.S. (2013i). Households' Access to Improved Drinking Water and Water Treatment Behaviour in Urban and Rural Nigeria *Asia Life Science*. Supplement 9: 283-300.
- Oyekale, A.S. (2014a). Ethnic Characterization of Gender-Inequality and Domestic Violence against Women of Reproductive Ages in Nigeria. *Studies in Tribes and Tribal*. 12(1): 31-41 (2014).
- Oyekale, A.S. (2014b). Impacts of Climate Change on Livestock Husbandry and Adaptation Options in the Arid Sahel Belt of West Africa: Evidence from a Baseline Survey *Asian Journal of Animal and Veterinary Advances*, 9(1): 13.

- Oyekale A.S. (2014c). Sex-Related Behaviour, HIV Knowledge and Incidence of Sexually Transmitted Infections (STIs) among Men of Reproductive Age in Cameroon, *Int. J. Environ. Res. Public Health*, 11, 12726-12738; doi:10.3390/ijerph111212726.
- Oyekale A.S. (2014d). Domestic Violence Against Women and HIV Vulnerability in Nigeria, *Mediterranean Journal of Social Science*, 5(20):190-201
- Oyekale, A.S. (2014e). Assessment of Pregnancy Status, Malaria Knowledge and Fever Morbidity among Women of Reproductive Age in Nigeria. *Iranian Journal of Public Health*, 43(9):1192-1203.
- Oyekale, A.S. (2014f), Factors Explaining Child Survival in Ethiopia: Application of Two-Stage Probit Model. *African Development Review*, 26: 237–249. doi: 10.1111/1467-8268.12078.
- Oyekale A.S. (2014g). Tribal Perspectives on Female Genital Mutilation (FGM) Among Married Women and HIV&AIDS Risks in Nigeria. *Studies in Ethno Medicine* 8(2):187-195.
- Oyekale A.S. (2015a). Assessment of Sustainable Land Management and Food Security among Climatic Shocks' Exposed African Farmers. *Disaster Advances*. 8(2):38-45.
- Oyekale AS (2015b). Climate change induced occupational stress and reported morbidity among cocoa farmers in South-Western Nigeria, *Annals of Agricultural and Environmental Medicine*. 22(2):361-365.
- Oyekale AS (2015c). Environment-Related Factors Influencing Morbidity among Economically Active Household Members in Urban and Rural Nigeria, *Environmental Economics*,
- Oyekale A.S. (2015d). Ownership of Mosquito Nets, Dwelling Characteristics and Mothers' Socio-Economic Factors Influencing Incidence of Malaria among Under-Five Children in Cameroon, *International Journal of Occupational Medicine and Environmental Health*, 28(3). DOI: 10.13075/ijomeh.1896.00340.
- Oyekale A.S. (2015e). Factors explaining access to and utilization of extreme climate forecast in sub-Saharan Africa (SSA). *Environmental Economics*. 6(1):91-103.
- Oyekale AS. (2015f). Assessment of Malawian Mothers' Malaria Knowledge, Healthcare Preferences and Timeliness of Seeking Fever Treatments for Children Under Five. *Int. J. Environ. Res. Public Health*, 12, 521-540; doi:10.3390/ijerph120100521.
- Oyekale AS (2015g). Access to Risk Mitigating Weather Forecasts and Changes in Farming Operations in East and West Africa: Evidence from a Baseline Survey, *Sustainability*, 7, 14599-14617; doi:10.3390/su71114599.

Oyekale AS (2015h). Factors Explaining Households' Cash Payment for Solid Waste Disposal and Recycling Behaviours in South Africa, *Sustainability*, 7:15882-15889.

Oyekale TO, Oyekale AS and Oyedepo JA (2015). Farm Households' Perceptions on Climate Change and Willingness to subscribe for Advisory Weather Forecasts in South West Nigeria. *Disaster Advances*, 8(11): 8-19.

Rasmussen, J. L. (1988). Evaluating outlier identification tests: Mahalanobis D Squared and Comrey D. *Multivariate Behavioral Research*, 23(2), 189-202.

Rasmussen, J. L. (1988). Evaluating outlier identification tests: Mahalanobis D Squared and Comrey D. *Multivariate Behavioral Research*, 23(2), 189-202.

Reiss PC and Wolak F. (2007). Structural econometric modeling: Rationales and examples from industrial Organization *Chapter 64, Handbook of Econometrics, Volume 6A Copyright © 2007 Elsevier B.V. DOI: 10.1016/S1573-4412(07)06064-3*

Roberto Pedace (2013). *Econometrics For Dummies*, ISBN: 978-1-118-53384-0, Paperback 360 pages.

Runge, C.F. (2006). "Agricultural Economics: A Brief Intellectual History," page 1 (abstract), University of Minnesota Working Paper WP06-1, June 2006, <http://ageconsearch.umn.edu/bitstream/13649/1/wp06-01.pdf>

Samuelson, P.A., Koopmans, T.C. and Stone, J.R.N. 1954. Report of the evaluative committee for *Econometrica*. *Econometrica* 22, 141–6.

Savoiu, G. and C. Manea (2013). A Continuing Approach, from Financial Economics to Financial Econometrics Or the Econometric Thinking Applied to Financial Economics, Romanian Statistical Review nr. 3 / 2013: 66-76.

Schiffler RE. Maximum Z Score and outliers. *The American Statistician*, Vol. 42, No.1

Schiffler, R.E. (1988). Maximum Z Score and Outliers, *The American Statistician* 42(1): 79-80.

Seo, S. (2002). A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets. Internet file retrieved on 15<sup>th</sup> March 2015 from <http://d-scholarship.pitt.edu/7948/1/Seo.pdf>

Shil, E. (1991). Remembering the University of Chicago, Teachers, Scientists and Students. University of Chicago Press.

Smith, A. 1776. An Inquiry into the Nature and Causes of the Wealth of Nations. Liberty Classics Edition (1981). Indianapolis, IN. Tukey, J. 1977. Exploratory Data Analysis, Addison-Wesley, pp. 43-44.

- Stevens, J.P. (1984). Outliers and Influential Points in Regression Analysis Psychological Bulletin, 95, 339 – 344
- Stock, J.H (2001). Macro-Econometrics. Journal of Econometrics 100:29-32.
- Taiwo, A.A. and A.S. Oyekale (2013). Urban Households' Willingness to Consume Vegetables Grown With Urine: A Case Study of Ibadan, Oyo State, Nigeria. *Asia Life Science*. Supplement 9: 173-184.
- Tejidini, S.A.R. and A.S. Oyekale. (2015) Factors Explaining Farm Households' Access to Climate Change Information in Ilorin West Local Government Area of Kwara State, Nigeria, *Environmental Economics*
- Tukey, J.W. (1977), Exploratory Data Analysis, Reading, MA: Addison-Wesley.
- Wainer, H. (1976). Robust statistics: A survey and some prescriptions. Journal of Educational Statistics, 1(4), 285-312.
- Wainer, H. (1976). Robust statistics: A survey and some prescriptions. Journal of Educational Statistics, 1(4), 285-312.
- White, H. (1980), A heteroskedastic-consistent covariance matrix estimator and a direct test of heteroskedasticity. *Econometrica*, 48, 817-838.
- White, H. (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, 48, 817-838.
- Zimmerman, D. W. (1994). A note on the influence of outliers on parametric and nonparametric tests. *Journal of General Psychology*, 121(4), 391-401.
- Zimmerman, D. W. (1995). Increasing the power of nonparametric tests by detecting and down weighting outliers. *Journal of Experimental Education*, 64(1), 71-78.
- Zimmerman, D. W. (1998). Invalidation of parametric and nonparametric statistical tests by concurrent violation of two assumptions. *Journal of Experimental Education*, 67(1), 55-68.