

METHODS OF QUANTITATIVE DATA ANALYSIS: WHEN AND WHERE APPROPRIATE

By

Dr. Kilishi A. Abdulhakeem

Department of Economics

University of Ilorin

meetkilishi@yahoo.com

kilishi-a@unilorin.edu.ng

08050221980

This version: March 2022

Department of Economics, University of Ilorin, Working paper series, No 06

Abstract

Data analysis is a crucial part of empirical research process yet many researcher do not have unhindered access to materials on quantitative methods for data analysis. This paper is an attempt at placing at the doorsteps of researchers (particularly nonexperts and early career academics) a broad range of quantitative methods for data analysis. The paper discusses some of the several methods for analysing quantitative data and when it is appropriate to use them. A researcher needs to understand when and why a method is appropriate to be used because employing a wrong method would yield wrong results and unreliable conclusions. The methods discussed were broadly grouped by their purpose, namely: descriptive statistics to understand pattern; ANOVA analysis to compare groups; correlation coefficients for measuring relationship; regression for measuring impact/effect; algebraic techniques for optimization and for evaluating the level of efficiency/effectiveness. The choice of technique of analysis from this robust list of methods is shown to depend on the unit of analysis, the way variables are measured and objective of study.

Keywords: Descriptive statistics, ANOVA analysis, Correlation analysis, regression analysis, Linear programming, Frontier analysis

1. Introduction

Data analysis is an integral part of empirical research. There are several methods and techniques to carry out data analysis. Choosing appropriate technique from the available numerous techniques may be challenging. Sometimes researchers use wrong method to analyse good data. Meanwhile a wrong choice of method would imply wrong analysis which would yield to wrong results and hence wrong conclusions as well as wrong policy advice. The choice of correct method largely depends on measurement of variables, unit of analysis and research objectives. Thus, it is important for every researcher to understand how to measure his\her variables correctly, employ the relevant/appropriately techniques/methods of analysis and interpret/ explain the results appropriately.

Therefore, this paper highlights the fundamental methods/techniques and modelling procedure that can be used to analyse different kinds of quantitative data. The methods are grouped according to the functions they perform. Specifically, different methods that can be used to achieve six separate research objectives are discussed. These include objectives to: understand pattern of a phenomenon; comparing different groups; measuring relationship between two variables; examining the impact of one or more variable(s) on another variable; investigating causality between variables; measuring effectiveness; and measuring efficiency.

Towards these ends, this paper is divided into four sections: section 1 is on introduction while section 2 discusses the differences qualitative and quantitative data. The various methods and models for analysing quantitative data are presented in Section 3, while section 4 concludes the paper.

2. Qualitative versus Quantitative Data:

Data can either be quantitative or qualitative. Quantitative data is a data set containing variables that can be expressed as a number or can be quantified by using numerical values. Examples of quantitative data include data on length, height, area, volume, weight, speed, time, temperature, humidity, sound levels, cost, ages, monthly income or saving of staff members of the faculty of social sciences, students' scores on examination, number of hours of study, etc. On the other hand, qualitative data are expressed using descriptive terms. Data that represent nominal scales such as

gender, socio economic status, and religious preferences are usually considered to be qualitative data.

The distinction between quantitative and qualitative data is based on the scale or level of measurement of the variable. Data can be measured in four different scales: Ratio Scale, Interval Scale, Ordinal Scale and Nominal Scale. Understanding how your variables are measured is very important because it determines the kind of technique to use.

Ratio scale: Variables measured in ratio scale have three characteristics: (i) the values of such variables can be arranged in ascending or descending order; (ii) comparing such as $X_2 \leq X_1$ or $X_2 \geq X_1$ is meaningful; (iii) calculating the ratio of the two variables such as X_2/X_1 is possible and meaningful; and the distance between the two can be meaningfully calculated. There are numerous examples of ratio data, salary of lectures, saving, national income, etc. We can arrange the salaries of staff of the faculty of social sciences from the smallest to the biggest or the other way. We can find the difference between the salary of a professor and associate professor or graduate assistant (GA). We can equally calculate the ratio of the salary of a professor to that of a GA.

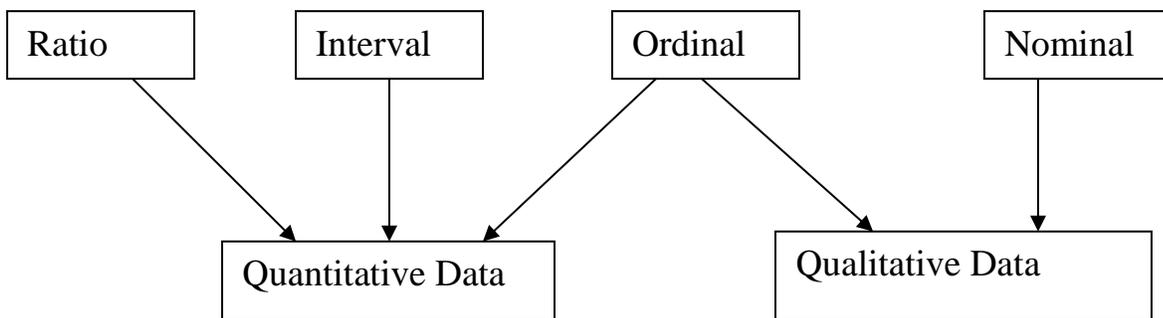
Interval scale: A variable measured on interval scale satisfies the first two of above-stated characteristics of data measured on ratio scale but lack the third feature of ratio data. The values of interval data are assumed to be equally spaced. Example is temperature on the Fahrenheit degree; different degrees of temperature can be arranged either in ascending or descending order, it make sense to compare level of temperature and say one is higher or lower but it does not make sense to calculate the ratio of temperature. In addition interval scale do not have true zero, for instance 0 degree of temperature does not mean absence of temperature. However, zero salary mean absence of salary.

Ordinal scale: A variable measured on ordinal scale satisfies only the first feature of above-stated characteristics of data measured on ratio scale. That is, its values can be ordered - can be arranged in either ascending or descending order. But the distance between its values cannot be meaningfully calculated. Example, even when we can rank the following options “Strongly Dislike”, “Moderately Dislike”, “Indifference”, “Moderately Like” and “Strongly Like” respectively as 1, 2, 3, 4 and 5, respectively, calculating the difference between “Strongly Dislike”

and “Strongly Like” as 5 minus 1 is not meaningful. Similarly, calculating the difference between one individual’s rating of “Strongly Dislike” and that of another person’s rating of “Strongly Like” as 5 minus 1 is not meaningful. It is straightforward to say that someone who strongly dislikes something dislikes it more than someone who just dislikes it. Therefore, it is possible to arrange ordinal data in either ascending or descending order as we did above. However, we cannot calculate by how much one dislikes it and we cannot also find the ratio between “Strongly Dislike” and “Dislike”. Another example is health status. Individuals can be classified as having very good, good, fair, poor, or very poor health status. Someone whose health is very good is healthier than one whose health is good, but it is difficult if not impossible to calculate by how much one person is healthier than the other or to find the ratio of health status of two individuals.

Nominal scale: A variable measured on nominal scale does not satisfy any of the above-stated characteristics of data measured on ratio scale. Thus, it can neither be arranged in ascending nor descending order; the difference between two values cannot be meaningfully calculated and their ratio cannot be meaningfully calculated. Examples are gender (male, female), marital status (married, unmarried, divorced), etc.

Figure 1: Categorization of Data into Quantitative or Qualitative Based on Measurement scale



Quantitative data are measured either in ratio scale or interval scale; however a few ordinal variables are categorized as quantitative data, e.g. rank of students in an examination scores. On the other hand qualitative data are measured in either ordinal scale or nominal scale. Another line of distinction between quantitative and qualitative data is that quantitative data are continuous variables while qualitative data are categorical variable. It is important to mention here that it is the scale of measurement not the source of the data that determines whether a variable is

quantitative or qualitative. There can be quantitative data from primary source or from secondary source. Similarly, qualitative data can be from primary source or from secondary source.

3. Methods of Data Analysis

There are several techniques that can be used to analyze quantitative data. However, the different techniques of methods can be broadly grouped according to the functions they performance or by the purpose/objective of research such as: (i) understanding the pattern of a phenomenon; (ii) comparing different groups; (iii) measuring relationship; (iv) measuring impact/effect/influence; (v) determining causality; (vi) evaluating the effect of an intervention; (vii) measuring efficiency. The various methods available for achieving each of these objectives are discussed in turns.

i. Understanding Pattern

There is gamut of methods to examine the pattern of phenomenon which can be discussed under the umbrella of descriptive statistics or summary statistics. Descriptive statistics comprises of useful tools for inspecting data which reveals the pattern, trend, and other information about the data in one view. There are several ways of inspecting data; the most common are the use of tables, graph and summary statistics. Tables display data with cases in rows and variables in columns. This will help the researcher to see how variables differ from case to case. Percentage and frequency distribution can also be displayed in tables; this will give a sense of which values occur most often. Graphs are essential tools for seeing the trend, how spread out the values are, and whether there are any values that appear to be unusually large or small (outliers). The most frequently used graphical tools are: Line graphs which facilitate comparisons of series; Scatter plots: give combinations of values from two series for the purpose of determining their relationship (if any); Histograms: give an indication of the distribution of the observations; Bar graphs; and Pie charts.

Descriptive statistics equally involves condensing data into a single number called a summary statistic. Some of the common summary statistics are mean; median; range; standard deviation; variance; skewness; kurtosis; and Jarque-Bera. Mean indicates the values around which all the values taken by the variable are distributed equally. Standard deviation is a measure of dispersion, mode is the value of the most often observed, median is the value such that the larger values are as many as the smaller values, skewness is zero for a symmetric distribution (i.e a normal distribution). Kurtosis is equal to 3 for a normal distribution, when the kurtosis is larger than 3 we

will say that the distribution has a fat tail. Jarque–Bera equally gives information about the distribution of data.

If a researcher is just interested in understanding the story inherent in a data set, descriptive statistics would be sufficient without necessarily conducting any rigorous analysis. However, for a researcher who is interested in further analysis, descriptive statistics will serve as preliminary analysis. For example, if the objective of a researcher is to know which among the six states in South West of Nigeria experienced more armed robbery cases on the average over a period of time. Computing the mean values of armed robbery over the time period for each state would provide sufficient information to achieve such objective.

For example, Figure 1 displays the average cases of armed robbery over ten year period for six states. It is easy to see at a glance that Lagos state had the highest cases of armed robbery, follow by Oyo state, while Osun state had the lowest number of armed robbery cases among the six states. Inter alia, the mean values cum the graph adequately tell the story about the armed robbery situation in the six states.

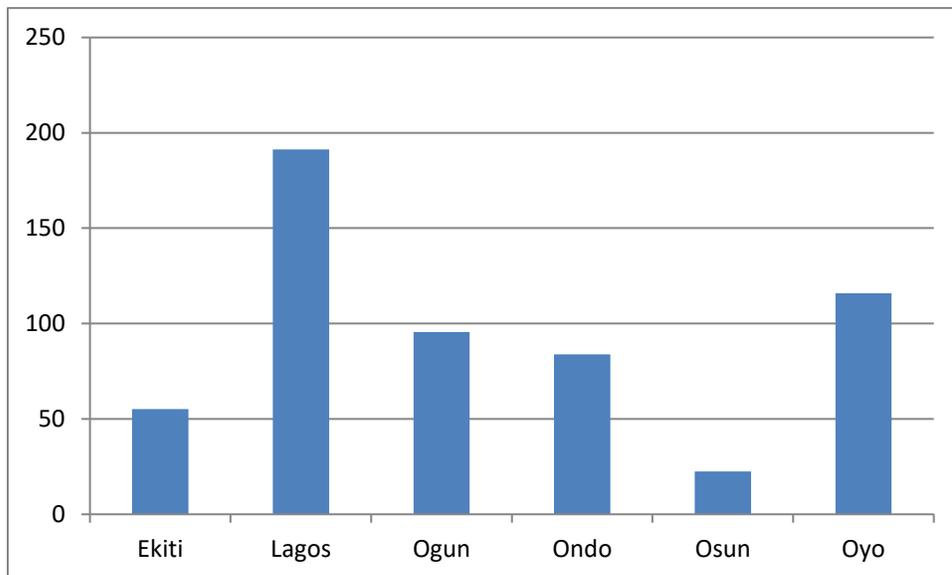


Figure 1: Mean Armed Robbery in Six States over Ten Years

ii. Comparing Groups

The objective of a research might be to compare two or more different groups. That is, the aim is to find out if one group is significantly different from another. For example, is the average academic performance of female significantly different from the performance of male? Is the profit of small businesses located in the urban areas significantly different from those located in sub-urban areas? Questions like these can be answered using t test, ANOVA analysis and MANOVA analysis among others depending on the number of groups. When the comparison is between two groups such the examples given above, t test is the most appropriate. But when the comparison among groups number than two, then ANOVA analysis will be suitable. MANOVA analysis is applicable when there are more than one variable of comparison and number of groups is also more than two. However, the variable(s) of comparison must be a quantitative (continuous) variable(s).

iii. Measuring Degree of Relationship

A researcher may be interested in investigating the existence, direction and strength of relationship between two quantitative variables. The measures of relationship do not in any way describe impact or direction of causality; they only verify existence of a relationship or association. Correlation coefficient is the most often used method to determine the existence or otherwise of linear relationship between two variables as well as to measure the strength of such relationship. The value of correlation coefficient ranges from 0 to 1; when it takes a value of 0 that means complete absence of linear relationship, while a value of 1 means perfect relationship. These values can either be positive or negative. Positive correlation coefficient means direct relationship and a negative value means inverse relationship. The closer the value of the correlation coefficient is to zero, the weaker the linear relationship and the closer it is to either -1 or 1, the stronger the linear relationship. Though, there are many different types of correlation coefficients, the most commonly used are Spearman (rank) and Pearson (simple) correlation coefficients. Pearson is more appropriate if the data are measured in interval or ratio scale while Spearman is preferred in case of ordinal data (see Erinoshio et al 2002; 78 and Mertler and Vannatta 2005; 8 – 9).

Suppose, a researcher computed correlation coefficient of the relationship between incidence of armed robbery and unemployment in Nigeria and the output is 0.65. If this value is statistically significant, then it implies that unemployment and armed robbery have direct and fairly strong

relationship. It worth noting that simple correlation coefficient is designed to detect the existence of linear relationship. If the relationship is non-linear correlation coefficient will result in a misleading conclusion. Suppose further, that the relationship between armed robbery are level of education gives correlation coefficient of -0.24 . This shows that there is weak inverse relationship between level education and armed robbery.

iv. Measuring Impact/Effect/Influence

Researchers are often interested in measuring the impact or effect of one or more variables (Independent Variables) on another variable called the dependent variable. In such situation, regression analysis is the most commonly and appropriate approach to use. Specifying a regression model correctly is very crucial in regression analysis because misspecification has some consequences on the validity/reliability of estimates. The first task in specifying a regression model is to identify the dependent variable and the independent variables and ensuring that relevant explanatory variables are included. The dependent variable is the variable that is explained while independent variables are variables that explain the behaviour of the dependent variable.

There are three important things that matter in choosing the independent variables to be included in a model. The first is theory which usually states the fundamental determinants of a variable. When there is an established theory, then it is mandatory to include the theoretical variable(s) even if they are outside the objectives of the research. For example, a researcher is interested in examining the impact of family size on household consumption expenditure. Given this objective, the dependent variable is household consumption expenditure, the independent variable is family size. However, a literature search will suggest that the major theories of consumption are absolute income hypothesis, permanent income hypothesis and life-cycle hypothesis. All tend to agree that income has some role to play in explaining the level of consumption expenditures. Hence, it will be wrong to estimate a model on consumption expenditure omitting income. Therefore, in addition to family size, the researcher is expected to include income as one of the independent variables.

The second thing to consider is empirical evidence which often provide guides on the additional explanatory variables to include. That is what are the variables found to be significant determinant of your dependent in previous empirical studies? Variable(s) that are not mention by theory but found to be significant in most previous studies have to be included. The third is common sense, this may provide guides on other explanatory variables to include in the model. That is, does

common sense support the inclusion of a variable as a regressor in the model? Some people may wonder why common sense is important in regression analysis. Preece (1987; 397) aptly puts it thus: “the procedures of good statistical practice are founded on experience and common sense; it is a good practice to stop and think before running a regression”.

There are three types of data: cross-section; time series and panel data. Cross sectional data is kind of data collected at a point in time across many units. The units could be individuals, household, firms, local governments, states or countries. On the other hand, time series data is the data gathered over a period of time on a given unit. Most often time series data is aggregated data, for example GDP of Nigeria from 1960 to 2020. And panel data is data that is collected across many units for over a period of time. For example, GDP of 15 West African countries over ten year period. Table 1 presents the main differences among the three types of data.

Table 1: Types of Data

Type of Data	Time Period	Number of Units
Cross section	One	Many
Time series	Many	One
Panel	More than one	More than one

There are different kinds of regression models and different methods of estimating them depending on the measurement of the variables involved particularly the dependent variable and the type of data. Table 2 below presents the commonly used regression models and methods. Basically, a variable can either be continuous or categorical. Continuous variable takes range of values, for example income in naira, age in years, profit in naira and so on. While categorical variable which also known as discrete variable take specific values. Discrete variables are of three types viz binary which takes just two values 1 or 0; ordered variable which takes more than two values and the values have natural ranking; and multinomial variable which equally takes more than two values but the values do not have natural ranking.

From Table 2, if a model involves a cross sectional data and the Dependent Variable (DV) is a continuous variable, the researcher can estimate such model using Ordinary Least Squares (OLS) while accounting for heteroskedasticity, or use either Generalized Least Squares (GLS) or Weighted Least Squares (WLS). However, if there is issue of endogeneity the researcher would have to use any of the Instrumental Variable (IV) estimators. Meanwhile, the DV is a discrete

variable either binary, ordered or multinomial, the model would be gauged by logit or probit modeling procedure. These two models use Maximum Likelihood Estimator (MLE). While logit uses MLE of logistic distribution, probit uses MLE of normal distribution.

In the case of time series data, the DV is most often a continuous variable. The characteristics of the variables involve determine the modeling procedure. There are number of pre-estimation tests which will reveal the characteristics of the variables that is necessary for the choice of appropriate model and method. If the variables are nonstationary and not cointegrated, the appropriate modeling procedure is first difference ARDL (Autoregressive Distributive Lag) for a single equation model and first difference VAR (Vector Autoregression) for a system of equations model. On the other hand, if there is cointegration, there are two options, either to estimate short run Error Correction Model (ECM) or to estimate long run cointegrating regression. The second option consists of three fully efficient estimation techniques, these are Fully Modified OLS (FMOLS), Dynamic OLS (DOLS) and Canonical Cointegrating Regression (CCR). Vector Error Correction Model (VECM) is applicable when series are modeled as system of equations and all series are I(1) as well as cointegrated.

Finally, if the DV in a static panel data model is a continuous variable, Fixed Effect (FE) Model, Random Effect (RE) Model, Population Average (PA), and Between Effect (BE) among others can be used to gauge the model. FE explores the relationship between predictor(s) and outcome variables within an entity (unit) such as household, firm, state, country. Hence, it is more useful when the interest is to analyse the impact of variables that vary over time. FE cannot be used to investigate the impact of time invariant variables on the dependent variable. However, FE removes the effect of time invariant characteristics so as to assess the net effect of the independent variables on the dependent variable. Since FE controls for all time invariant differences between the individuals, the estimated coefficients cannot be biased because of omitted time invariant characteristics. FE does not work well with data for which the within variation is minimal, or for slow changing variables over time. It will poorly identify the coefficient of variable with small within standard deviation. On the other hand, RE assumes that the variation across units are random and uncorrelated with the independent variables. RE is useful when differences across units have some influence on the dependent variable. Therefore, RE can estimate the impact of time invariant variables. RE essentially uses GLS estimator, and there are different versions of this estimator. Meanwhile, in case the DV is discrete, logit and probit will be the most appropriate estimation techniques.

However, for dynamic panel model, FE, RE, PA, BE and others will not be appropriate because of the problem of endogeneity caused by the introduction of lag of DV as one of the predictors. The most commonly use techniques are the First Difference GMM developed by Arellano and Bond, or the System GMM by Blundell and Bond on one hand and Arellano and Bover on the other hand. The System GMM is an advancement over the First Difference GMM.

Table 2: Different Models and Methods for Regression Analysis

Measurement of DV	Type of Data	Model/ Method	Remark
Continuous	Cross-section	OLS, GLS, WLS, IV, 2SLS etc	Always use robust option when using OLS; IV & 2SLS are applicable when there is endogeneity in the model
Continuous	Time-series	OLS, ARDL, ECM, DOLS, CCR, FMOLS, VAR, SVAR, VECM,	See Kilishi A. A. (2022) Modeling with Time Series: Issues and Common Errors, for details on selecting appropriate method
Continuous	Panel Data	FE, RE, PA, BE, GMM, PMG, MG, DFE, Hausman-Taylor, Amemiya-MaCurdy	There are several versions of RE; FE has few variants; GMM is for dynamic panel; PMG, MG & DFE are for non-stationary panel; last two are used when there is endogeneity
Binary or Dichotomous	Cross-section panel data	Logit, Probit	The estimator for both is essentially MLE
Ordered or ranked	Cross-section panel data	Logit, Probit,	
Multinomial	Cross-section panel data	Logit, Probit,	

v. Determining Direction of Causality

Sometime a researcher may want to know whether there is causality between two variables and what is the direction of the causality. For instance, is it x that causes y or is it y that causes x or that both x and y cause each other simultaneously. In such a situation, causality analysis would have to be undertaken. A well known framework for providing answers to such questions is Granger Causality test framework. There are different variants of this procedure today which can be used to handle issues of causality in panel data models.

vi. Measuring Effectiveness

Techniques for measuring efficiency and effectiveness are somehow different from statistical tools of data analysis. Most of these techniques are non-parametric, mathematical and algebraic in nature. The most common techniques in this category are Linear Programming (LP), Goal Programming (GP), and Game Theory among others. The first two are mathematical programming methods that are used in determining the most effective/optimal action or decision. The prominent areas of application of these techniques in social and management sciences include: production management such as production planning, assembly–line balancing, and blending problems; financial management such as portfolio selection, profit planning, and risk minimization; marketing management such as physical distribution, selection of advertising media, and travelling salesman problem; and personnel management such as optimum allocation of manpower, determination of equitable salaries and job evaluation and selection. The main difference between Linear Programming (LP) and Goal Programming (GP) is that while GP is used in case of multi-objective problem, LP is an approach for solving single objective. Some of the computer software that can be used to estimate LP and GP models are STORM, TORA, QSB, LINDO, MIQP, MILP, CBC, among many others.

Game theory provides a series of mathematical models that may be useful in explaining interactive decision making concepts, where two or more competitors are involved under conditions of conflict and competition. A good example of situation when Game Theory is very useful is pricing of product particularly where sale of product is determined not only by its price but also by the price set by other producers (competitors) of similar product. GAMBIT is an open – source computer software that can be used for solving game theory type of problems.

vii. Measuring Efficiency

Stochastic Frontier Analysis (SFA) and Data Enveloping Analysis (DEA) are extensions of Linear Programming technique that provide means of calculating apparent efficiency levels within a group of organizations. The efficiency of an organization is calculated relative to the group's observed best practice. There are many types of efficiencies but the most important ones are technical efficiency and allocative efficiency.

SFA and DEA are powerful tools of gauging relative efficiency estimates of different Decision Making Units (DMU) at the same time. Thus one can identify the units that are efficient and those

who are inefficient. Since the efficiency estimates can be related to individual unit's characteristics, the source of inefficiency can therefore be identified. For example, SFA and DEA can be used to answer questions like has efficiency of banks changed after deregulation? Has this change varied across different banks? This may be more difficult to achieve using conventional econometrics because deviations from efficient/optimal path are modeled as statistical noise in conventional econometrics. However, every producer may attempt to optimize, it is most likely that not all of them may succeed in their efforts. For example, given the same inputs, and the same technology, some will produce more output than others, i.e., some producers will be more efficient than others. Hence, SFA and DEA have advantage over conventional econometrics in this regard.

These two techniques, SFA and DEA are supposedly estimating the same underlying efficiency values but the nature of the two methods are very different. To a large extent, SFA and DEA are competing techniques. No formulation has yet been devised that unifies the two in a single analytical framework. Arguably, the former is a fully parameterized model whereas the latter is nonparametric but deterministic albeit also a-theoretical in nature. Nevertheless, the two have been extensively applied in performance evaluation and benchmarking of schools, hospitals, bank branches, production plants, and studies on cost, revenue, profit and other models of goal attainment. According to Kooreman (1994) the difference between the two techniques "is partly reflected by the types of information used by the two methods. DEA uses information on input and output quantities, and as such it can only address the issue of technical efficiency. SFA uses information on input prices, output quantities and total costs, and the inefficiency measured by SFA may be a combination of technical and allocative efficiency. Without further assumptions, the SFA method is unable to separate the two sources of inefficiency. The distinction between allocative and technical efficiency is important because in general they require different policy responses.

A number of computer software can be used to estimate SFA and DEA models; these include Frontier, LIMDEP, NLOGIT, DEAP, DPIN and TFPIP among others. The STATA and R software can also be used to estimate efficiency frontier models. It is important to mention here that there are many open source software for efficiency analysis.

4. Concluding Remarks

Quantitative data analysis is a broad field in research methodology. This paper discusses what is quantitative data and quantitative data analysis. It outlines the major differences between qualitative and quantitative data as well as discusses some important techniques for analyzing quantitative data. Tables, graphs and summary statistics are shown to be useful tools for descriptive statistical analysis of quantitative data. Other more robust and precise techniques for analyzing quantitative data were also discussed; including correlation analysis, ANOVA analysis, regression analysis, linear programming, goal programming, game theory, stochastic frontier analysis and data enveloping analysis, among others.

It concluded that the choice among all these methods depends on the problem at hand or objective of the study, the level of measurement of variables and unit of analysis. A researcher might have three to four or more objective in single research work that will require him/her using different method or technique to achieve each of the objectives. This paper is just a guide to help researchers identify the methods or techniques that are suitable for different kind of quantitative data analysis. However, it is not possible for all the available methods and techniques to be discussed in a single paper, thus, the paper focused on the fundamental ones.

Bibliography

- Baltagi H. B. (2005), *Econometric Analysis of Panel Data*, West Sussex: John Wiley & Sons, Inc
- Donnelly A. R. (2007), *Then Complete Idiot's Guide to Statistics*, New York: Penguin Group, Inc.
- Enders W. (2010), *Applied Econometric Time Series*, Hoboken: John Wiley & Sons, Inc
- Erinosho L., Obasi N. I. and A. Maduekwe, (2002), *Interdisciplinary Methodologies in the Social Sciences*, Abuja: AUSCON FIRESEED& Co. LTD
- Greene H. W. (2008), *Econometric Analysis*, Upper Saddle River, New Jersey: Pearson Prentice Hall
- Gujarati N. D. and D. C. Porter, (2009), *Basic Econometrics*, New York: McGraw Hill
- Hamilton D. J. (1994), *Time Series Analysis*, Princeton: Princeton University Press
- Hill C. R., Griffiths E. W. and G. C. Lim, (2012), *Principles of Econometrics*, Hoboken: John Wiley & Sons, Inc
- Kooreman P. (1994), Data Envelopment Analysis and Parametric Frontier Estimation: Complementary Tools, *Journal of Health Economics*, vol 13, pp 345 – 346

- Maddala S. G. (1992), *Introduction to Econometrics*, New York: Macmillan, Inc
- Mertler A. C. and R. A. Vennatta, (2005), *Advanced and Multivariate Statistical Methods: Practical Application and Interpretation*, Glendale California: Pyczak Publishing
- Preece D. A. (1987), Good Statistical Practice, *The Statistical*, vol.36, pp 397 – 408
- Sharma K. J. (2013), *Operations Research: Theory and Applications*, Delhi: Macmillan Publishers India Ltd
- Wooldridge M. J. (2002), *Econometrics Analysis of Cross-Section and Panel Data*, Cambridge Massachusetts: MIT Press