Volume 1, Issue 1, pp 27-34; March 2012

Online Journal of Social Sciences Research

Full Length Research Article

Available Online at http://www.onlineresearchjournals.org/JSS

The Challenges of Applied Econometric Research

Bruno R .M. Ocaya

Chair, Department of Statistics and Actuarial Science, Makerere University. Email: bocaya@isae.mak.ac.ug.

Downloaded 6 February, 2012 **Accepted 9 March, 2012**

This paper addresses three key issues regarding the challenges of applied econometric research. First, it highlights major challenges researchers often confront when applying econometric tools for empirical analysis. These challenges are analyzed around problems and difficulties associated with data, model building and specification; and peculiarity of computing software. Because of their unique data challenges, the appendix to this paper discusses pertinent data problems in Third World countries and Sub-Sahara Africa in particular.

Secondly, the paper makes suggestions and recommendations in addressing and tackling some of the challenges analyzed. While the paper makes proposals for handling Third world data problems for econometric research, it recognizes that Africa still needs to build a strong statistical data base through its commitment to develop and strengthen statistical institutions in the region.

Finally and most importantly, the paper provides guidelines of major issues for the realization of plausible empirical econometric research. They include among others the following: (i) Good knowledge of data for appropriate action to address their deficiencies, (ii) Sufficient data in order to lend credence to estimates, tests of hypotheses and confidence intervals, (iii) Parsimony as a tenet for building models which address pertinent research questions, (iv) Familiarity with econometric techniques/ models and statistical computing packages; and (v) Exploration of alternative forms of data and models for better results and interpretations.

Keywords: applied econometric research; data challenges/problems; model building; econometric modeling and computing; methodology.

INTRODUCTION

Though mathematically and statistically challenging, theoretical econometrics is elegant and straight forward. It is void of real life data complications, challenges and problems. Most data used for theoretical econometrics are trumped-up or experimentally generated/simulated for illustrative purposes[1]. Similarly, there are no major challenges in the computations and in the building and specification of theoretical models since they are meant to mimic rather than replicate real data situations [1].

Nevertheless, the nature of data encountered in applied econometric research leaves a lot to be desired. The paucity and reliability of data in particular pose major challenges to empirical work especially in Third World countries where the quality of data is generally still very poor and measurement perceptions of variables are varied and weak [2]. Likewise, the choice and specification of appropriate models (i.e., model building) are often difficult tasks involving complex empirical issues. Besides the nuances and peculiarity of statistical computer software, results of computations from empirical econometrics usually bewilder and frustrate researchers.

The motivation of this paper is to highlight and analyze major problems, difficulties and challenges econometricians often tend to confront in their applied research. It will also attempt to advocate for possible solutions to address and tackle them. Before delving into these issues, we will for completeness first analyze basic useful concepts of data for empirical research.

Review of Data types, Sources and Collection

Data comprises of disaggregated and aggregated data which are respectively used for micro and macro analyses. The sources of disaggregated data are the

[1] An example of such data is those obtained through various Monte Carlo type studies.

micro units comprising of individuals, households and firms among others. The aggregation of micro units' data yield macro data. The macro data are usually compiled by governments, financial and international agencies among others [3].

There is also primary and secondary data which describe the sources by which data is collected. Strictly speaking, primary data is collected at the initial source based on a designed questionnaire. Secondary data are those obtained or compiled from other sources. The main sources of secondary data are reports and publications made by various agencies. Data from secondary sources are also available on the respective web pages in the internet for easy downloading[2].

Data can be observed or collected at a point in time. Such data are referred to as cross -sectional data. Time series data are those which are observed or collected over a period (annual, semi-annual, monthly, quarterly, weekly, daily, etc.). A combination of cross-sectional and time series data gives rise to what is known as panel data or pooled/longitudinal data [4].

Data collected or compiled can be qualitative or quantitative. The type and purpose of data including method of collection and computation are different and vary depending on the interest of bodies that compile them (i.e., governments, organizations, agencies, individuals, etc.).

DATA PROBLEMS

Measurement, Accuracy and Disparity of Data

The definition and measurement of variables are formidable problems of data. Because several definitions are attached to a variable, measurement of many variables has in general been a controversial issue in most empirical and econometric studies, even when theoretical foundations are in place [5]. Where standard procedures for measurement exist, issues of data accuracy are raised especially when data collection and compilation are from different sources.

Data accuracy may be influenced by biases motivated by personal, political or institutional interests. For empirical analysis, data accuracy is not as important as the consistency of measuring data from a given source, more so if a data source forms the basis of policy and is broadly considered (understood) to be representative[3]. Nonetheless, bad data base on strong hypothesis or on weak theoretical foundation are no good for empirical analysis [6].

Disparities in data often emanate from varied perceptions, which have given different meanings in measuring variables in different countries. This is particularly the case with qualitative variables where quantification is problematic partly because well-defined acceptable norms of measurement do not yet exist. Even where measurement problems do not arise, different methods of data analysis have led to different conclusions [7].

Although rebuttals clamor for improvement in data collection, it is generally recommended that empirical studies be carried out mainly in areas where data are readily available and measurement problems are minimal. Where measurement is not possible, it is permissible to use data of proxy variables. Similarly, definitions and measurements of variables should be clear and concise for model building (i.e., units of measurement, proxy/instrument variables used and transformations carried out) should be clearly spelt out for correct data interpretation and computation[4]. Excessive transformation of variables to answer theoretical requirements should be minimized or avoided altogether.

Missing, Unavailable and Insufficient Data

For various reasons, data for conducting meaningful econometrics research can be undermined and constrained by missing, insufficient and even unavailable information from different sources. Such problems are very common with time series data particularly those for Third world countries. However, problems of cross sectional data are normally resolved within the realm of statistical sampling [7].

Missing values of a variable are normally replaced with their means. This standard approach presupposes absence of secondary sources, and is only plausible when there are relatively few missing values in a given series of a variable. Although variables with few missing values are uncommon, the following guidelines may be adopted to handle missing values of variables:

(a) Use secondary sources to obtain information (clues) which would lead to reasonable estimates of the missing variables. Secondary sources are also useful for cross checking errant values (outliers). If the source of the error is difficult to ascertain, replace outliers with mean values or discard them if they are few.
(b) Apply known (established) relationships to approximate missing data for your variables.
(c) Discard the variable when (a) or (b) are not possible or when it is impossible to determine missing values for the variable.
(d) When most variables have many missing values

[2] Usually such data are downloaded and opened into a spreadsheet for analysis, manipulation and computation.
[3] For example the Ugandan economy has for a long time been quoted and taken for granted to grow at the rate of 8 % per annum because it is trumped-up by specific sources.

[4] For example, variables expressed in real terms should specify the base year.

which cannot be obtained from any other source, it is considered unavailable. Data for such variables should be dropped from the study altogether on grounds of unavailable information.

Data points with less than 30 cross sectional observations are generally regarded insufficient (inadequate) for empirical research. It is generally recommended that one should have at least 30 data points on cross sectional data and about 50 for time series data. Insufficient data points do not ascertain normality and therefore can undermine hypothesis tests and construction of confidence intervals [8]. In the event that no additional points can be obtained, researchers have opted for interpolation of data series. Several interpolation methods for converting data to different frequencies has been proposed and most statistical packages these days convert them automatically[5]. It should be noted however that interpolation methods are only suggestive and do not provide true values of the underlying series.

Unlike data on time series, data on cross sectional variables are usually available for primary data. Similarly, although there are shortcomings of questionnaire design and follow-ups, the problems of cross-sectional primary data are generally limited because units or variables with missing, unavailable or insufficient data can be discarded or added as the case may be.

Data Manipulation and Transformation

Data from various sources may not necessarily be given in the way it is desired for analysis and computation. They have to be therefore converted to desired forms through manipulations[6] and appropriate transformations. For time series economic variables, it may entail: changing nominal values to real values, changing data to a different base year, calculating growth rates and merging data published from various sources with different base years for consistency, etc. Although various countries/institutions have different approaches and methods for data transformation, below are suggestions you might find useful in data cleaning:

Nominal versus real values

In order to adjust for differences in purchasing power through time and over space, nominal values have to be converted into real values through the formula:

$$\text{Real } X_t = \frac{\text{Nominal } X_t}{PI_t} \times 100 \qquad (1)$$

where X_t is any dollar dominated variable at time t, and PI_t is an appropriate price index at time t. The term “an appropriate price index” implies not only the choice of an index to use but also the adoption of an index expressed in a desired base year. There are several indexes used for converting various nominal data to their real values. The most frequently used are the Consumer Price Index (CPI), the Implicit GDP deflator and Producer Price Index (PPI).

Changing data to a different base year

Data from most sources are reported with a different base year. Some of these have to be changed to suit the purpose of the investigation either because the base year is too remote or does not capture the event of importance. Changing the base year of an annual data involves dividing the price index of each year by the price index of the desired new base year, then multiplying the ratio by 100

i.e.

$$PI_t^* = \frac{PI_t}{PI_{t,0}} \times 100 \qquad (2)$$

where

PI_t^* = the new price index at time t

PI_t = the price index at time t

$PI_{t,0}$ = the price index of the desired new base period

After generating a new series of price index PI_t^*, variables are converted to their real values corresponding to the new base year as previously discussed. It is important to note that such a transformation changes the base of index, but does not change the weights.

Related to the change of base year are cases where one source reports earlier observations of continuous data in one base year while another source reports more recent observations with a different base year. The ratios of overlapping observations from both sources are used to merge the data and obtain an index series consistent with the recent base year. For a one period overlapping observation, the ratio of the price index at the overlapping point can be used to extrapolate the index series backwards under the assumption that this ratio will pertain to all earlier observations[7]. This ratio is

[5] Usually we raise more observations by converting data from a lower frequency such as annual data to higher frequencies like monthly, quarterly, etc,.

[6] Manipulation of data may also involve arrangement of data in an appropriate format for the computing program (software) to read and correctly interpret.

[7] This formula may formally be stated as

weighted where there are several overlapping observations. The new index series of the recent base year is used to convert variables to their real values.

Whether time series, cross sectional or panel data, variables expressed in nominal or real values may produce unsatisfactory empirical results. Under such circumstances it is advisable to transform variables as ratios, percentages and growth rates or adopt, where possible, appropriate variable indexes or replace them with their proxies.

Growth measurements

Different measurements of growth rates can be determined accordingly. The discrete growth rate for a one-period annual data for variable X is derived from the formula[8]

$$G_x = \frac{(X_t - X_{t-1})}{X_{t-1}} \times 100 \text{ Or } G_x = \left[\left(\frac{X_t}{X_{t-1}}\right) - 1\right] * 100$$

$$PI_t(t_2 = 100) = \frac{PI_t(t_2 = t^*)}{PI_t(t_1 = t^*)} \times PI_t(t_1 = 100)$$

where

$PI_t(t_2 = 100)$ = the price index of recent observation for the recent base year.

$PI_t(t_1 = 100)$ = the price index of early observation for the early base year.

$PI_t(t_2 = t^*)$ = the price index of recent observation at time t.

$PI_t(t_1 = t^*)$ = the price index of early observation at a time t*.

$\frac{PI_t(t_2 = t^*)}{PI_t(t_1 = t^*)}$ = the ratio of the price index of recent to early observations at a time t*.

t* = the overlapping point.

[8] The computations of annual growth rate vary depending on the frequency of the data. For monthly and quarterly data, the rates are given respectively as

$$G_x = \left[\left(\frac{X_t}{X_{t-12}}\right)\right] - 1] \times 100 \text{ and}$$

$$G_x = \left[\left(\frac{X_t}{X_{t-4}}\right) - 1\right] \times 100.$$

The compound annual growth rates for these data frequencies are determined from the following formulas:

$$G_x = \left[\left(\frac{X_t}{X_{t-12}}\right)^{12} - 1\right] \times 100 \text{ (Monthly data) and}$$

$$G_x = \left[\left(\frac{X_t}{X_{t-4}}\right)^{4} - 1\right] \times 100 \text{ (Quarterly data).}$$

The annual "change" in percentage growth of variable X can be calculated from the absolute change in annual percentage growth

i.e.

$$\Delta G_x = \left[\frac{X_t - X_{t-1}}{X_{t-1}} - \frac{X_{t-1} - X_{t-2}}{X_{t-2}}\right] \times 100 \Rightarrow \left[\frac{X_{t-2}X_t - X_{t-1}{}^2}{X_{t-1}X_{t-2}}\right] \times 100 \quad (3)$$

The absolute change in X may also be obtained by estimating the following linear regression model

$$X = b_0 + b_1 T + \varepsilon$$

where T is the unit time interval and $\hat{b}_1$ is the absolute change in X over a given period T.

A one period instantaneous rate of growth G_x is approximated from the exponential formula

$$X_t = e^{G_x} X_{t-1} \Leftrightarrow \ln X_t = G_x \ln X_{t-1} \Rightarrow G_x = \ln\left(\frac{X_t}{X_{t-1}}\right) \times 100.$$

An alternate approach to computing instantaneous growth rate of variable X is by estimating the following log-lin regression model

$$\ln X = b_0 + b_1 T + \varepsilon$$

where T is the time trend and $b_{1.}$ becomes the instantaneous rate of growth.

The relationship between nominal and real growth rates is useful in obtaining lacking information especially where the percentage changes in real and nominal values are small. For a dollar denominated variable X, this relationship is approximated by[9]

[9]Akin to this formula is that for estimating non-dollar dominated magnitudes such as interest rates which represent costs of lending and borrowing money. For a given period, the real interest rate $\mathbf{I}_R$ is given by

$$\mathbf{I}_R = \mathbf{I}_N - \mathbf{\Pi}^e, \ \infty \le \mathbf{I}_R \le \infty$$

where $\mathbf{I}_N$ is the stated nominal interest rate, and $\mathbf{\Pi}^e$ is the expected rate of inflation. The snag in actualizing this relationship lies in the measurement of $\mathbf{\Pi}^e$. Thus far, suggestions range from using surveys of respondents to adoption of proxy variables. The construction of proxy variables would require theoretic assumptions about how people

$\% \Delta$ in Real X is approx. equal to $\% \Delta$ in Nominal X minus $\% \Delta$ in PI (Inflation rate)[10].

PROBLEMS OF ECONOMETRIC MODELING AND COMPUTING

Theoretical econometrics is elegant, neat and straightforward; devoid of major data problems and challenges. It employs the knowledge of mathematics and statistics to develop techniques and theorems for analyzing `random data based' empirical problems. The challenges of empirical econonometrics in applying these theoretical techniques and theorems can be surmountable given the problems of data and difficulties encountered in the specification and building of models [9]. For example, the adoption of inappropriate models and preoccupation with data and model perfections often lead to inaccurate interpretations of results, attainment of bewildering results and results which fall far short of expectations. The challenges and frustrations of results from applied econometrics can unfortunately lead to data mining- "panel beating" of data to confess desired outcomes *a priori* [10]. Problems of econometric modeling are ingrained in the difficulties and challenges of model building while those for computing lie in the appreciation of various computer packages to ensure correct computations.

Model Building

Model building is an "art" requiring wide-ranging knowledge and judgment which can be perfected with practice overtime[11]. It is also a "science" which applies the skills and knowledge of mathematics, statistics and computer science to develop appropriate techniques and theorems for analyzing empirical data [9]. An accomplished econometrician is one grounded in mathematics, statistics and computer science for empirical analysis. The application of econometric methods and models are no longer confined within the realm of economics. Today they transcend many disciplines in the sciences and social sciences including medicine.

Econometricians build models for two main reasons:

(i) In order to make inferences: This involves estimation of parameters (coefficients); hypothesis testing of parameters to establish their statistical significance and their theoretical validity; and construction of confidence intervals. Hence, conditions and data requirements that do not undermine statistical inferences should be seriously considered.
(ii) In order to undertake predictions (forecasts) which help in the formulation of policies: Econometricians (economists) analyze models and predict outcomes which act as a basis for guiding polices in achieving certain outcomes and policies for avoiding unfavorable ones. Therefore, simplistic parsimonious models which predict and capture silent features of an economic phenomenon are often encouraged.

Depending on the purpose of the research, models may be judged on "*a priori* theoretical (economic) grounds" or on their strength in analyzing and predicting the economic problem at hand. The main features of good econometric modeling are:

(i) Proper grasp of the research problem and knowledge of institutional realities of the environment at hand, theories being investigated, etc.

(ii) Identification of appropriate variables for the model including transformations deemed necessary.

(iii) Appropriate model selection which addresses pertinent research questions or hypotheses under investigation. These among others involve the following choices:

● Structural econometrics models vrs time series econometrics models.

● Single equation vrs simultaneous equations models.

● Models of different functional forms.

Note that model specification entails sound knowledge of econometrics and acquaintance with various econometric models. Data problems and model specification are inextricably linked in the sense that data limitations may constrain the appropriate (feasible) model specification and, conversely, model specification will have strong implications for the data required [11].

A useful model therefore should be "parsimonious, plausible and informative". Since model building is an art as well as a science, a practical researcher should not be bewildered by theoretical niceties and diagnostic tools of econometrics. Peter Kennedy [12] aptly advocated for the following "*Ten Commandments*" as a precaution to model building for the applied econometrician:

i. Thou shalt use common sense and (economic) theory.
ii. Thou shalt ask the right questions (i.e., put relevance before mathematical elegance).

---

form inflationary expectations. Researchers usually circumvent the problem by appealing to theories of adaptive and rational expectations.

[10] This formula holds exactly for exponential growth rates.
[11] In most empirical work, model building and computation is not a one time task achievement but rather a product which goes through a sequence of trials before an appropriate model is found.

iii. Thou shalt know the context (do not perform ignorant statistical analysis).

iv. Thou shalt inspect the data (e.g. undertake quick descriptive statistics to appreciate your data).

v. Thou shalt not worship complexity. Use the KISS principle, that is, keep it short and simple or principle of parsimony.

vi. Thou shalt look long and hard at thy results. (Never ever expect perfect results from data).

vii. Thou shalt beware the costs of data mining (i.e. beat the data into submission to obtain the desired conclusion).

viii. Thou shalt be willing to compromise (do not worship textbook prescriptions).

ix. Thou shalt not confuse significance with substance (do not confuse statistical significance with practical significance).

x. Thou shalt confess in the presence of sensitivity (that is, anticipate criticism).

Computing in Econometrics

There are a number of computer software available for computations and analysis of data in econometrics. The popular packages are the statistical and spreadsheet software. Most of these programs (software) are now ``menu driven'' and ``user friendly''. Although they vary in size, complexity, cost and programming requirements, almost all can do the necessary econometric computations. However, some of the programs are specialized and others require manuals to be able to carry out sophisticated calculations.

The main statistical packages currently in vogue are: E-views (time series), Rats (time series), Limdep (limited dependent variables), Stata (cross-sectional/general), Shazam (general) and SPSS which is mainly for social scientists. These statistical packages are driven by unique commands (codes) which interpret data to execute computations. It is absolutely essential for one to know how a particular software interprets and executes commands (codes) for the realization of correct econometric results. To safeguard against the execution of incorrect and inappropriate commands, it is always prudent to use text book data examples for comparison of your computations.

In addition to the statistical programs, there are spreadsheet software which econometricians use mainly for data analysis (i.e., data entry, coding, manipulation, etc)[12]. Usually, cross-sectional/time-series or panel data are entered as columns or as rows in a spreadsheet with names of variables in the first column or first row. Excel and Lotus are currently the most widely used spreadsheets packages. Spreadsheet software are also very useful for downloading/opening of data from various sources including the internet. Depending on software conformability, data from spreadsheet software are imported into statistical software for computation. Statistical packages on the other hand mainly use mathematical formulas for computations, data transformation and data generation.

There are also a number of mathematical software that one may find useful for undertaking basic econometric data analysis and computations. One particular software I would recommend for econometrics work is "Scientific Workplace". This software (combines *tex* and *latex*) has an inbuilt ``maple engine'' convenient for carrying out mathematical computations and simple statistical calculations encountered in econometrics. Its word processor enables one to write elegant complicated scientific formulas and notations found in econometrics text books.

Many users have different preferences for certain software for a number of reasons (e.g., acclimatization, training, discipline or degree of ``user friendliness''). Although it is advisable to use a manual and later versions of a particular software (because it incorporates new features), it is not absolutely necessary to learn software for their sake.

CONCLUSION

It is not feasible to address comprehensively all the challenges of applied econometrics that may arise in a wide variety of research studies that one may confront. Nonetheless, based on the above analysis, I wish to conclude by recommending consideration of the following major issues when undertaking applied econometric research.

First and foremost one should understand the nature of data in order to address their deficiencies appropriately. In particular, precaution should be taken in appreciating the units of measurement and any transformations that may have been made on data. A snapshot of descriptive statistics including graphs computed from statistical software is usually a good starting point in comprehending data [13]. Similarly, one should ensure that there is reasonably adequate data which theoretically justifies its use and gives credibility to estimation, tests of hypotheses and confidence intervals. The challenge of insufficient data points is a time series data problem which is common to data from most third world countries. It is a caveat which should be addressed upfront before considering building an econometric model. When adequate data is not forthcoming, one may, as a last resort, opt for the controversial method of data conversion to get more data points from a low frequency time series data [11].

Second, it is important to clearly specify your research question(s) in order to adopt the appropriate econometric model(s) which address them. Note that various econometric models address different research questions and investigate different hypotheses. Do not use or impose data onto models which have nothing to do with the research questions you want to investigate. When building an adopted model, identify its salient features adhering closely to the principle of parsimony. Likewise, take note of warnings which are pertinent to building your model that may have been raised by Peter Kennedy in his "Ten Commandments of Applied Econometrics"[2].

Third, it is advisable that one should be conversant with at least one statistical package and one spreadsheet software in order to be able to undertake required econometric computations [4]. The researcher should be able to clearly spell out hypotheses of the model based on better understanding of the nature and source of data although conventional tests (rule of thumb) stipulate that we *reject the null* at 5% and 1% significance levels when p-values are lower than 0.05 and 0.01 respectively [4]. Sometimes econometric better results may require exploration of other different functional forms of the model. For example, transformation of variables into logs or ratios/percentages can improve the fit of a model[12]. However, unnecessary transformation may also create difficulties in the interpretation of results.

Finally, Models based on purely time series data are prone to problems of *autocorrelation* (serial correlation) while those for cross-section data are susceptible to problems of *heteroscedasticity*. The researcher should make sure that these problems are addressed accordingly if they are serious. Likewise, It should also be noted that econometric models based on purely time series data tend produce *spurious* regressions results and undermine conventional test hypotheses and forecasting [4]. To circumvent the problem, ensure that the variables in the regression are *cointegrated* or the data series for the variables are *stationary*[13].

References

[1]. Hendry DF. Econometrics – alchemy or science? Econometrics 2000; 47: 387-406.

[2]. Mukherjee C, White H, Wuyrs M. Econometrics and Data Analysis for Developing Countries. London: Routledge 1998.

[3]. United Nations. Improving social statistics in developing countries: Conceptual framework and methods, Washington D.C. 1979.

[4]. Greene W. Econometric analysis, 6th ed., Englewood Cliffs, N.J.: Prentice Hall. 2003.

[5] Scott C. Measuring up to the measurement problems: The role of statistics in evidence-based Policy Making, London School of Economics, UK. 2005.

[6]. Grilliches Z. Data and econometrics – the uneasy alliance. Am. Econ. Rev. Papers Proceed. 1985; 75: 196-200.

[7]. Backhouse R, Morgan M. Introduction in data mining a methodological problem? J. Econ. Method. 2000; 7: 171-181.

[8]. Driscoll M. The ten commandments of statistical inference. Am. Math. Monthly 1977; 84: 628.

[9]. Bailar B. Statistical practice and research: the essential interactions. J. Am. Statistical Associat. 1988; 83: 1-8.

[10]. Hand D, Blunt G, Kelly M, Adams N. Data mining for fun and profit. Statistical Sci. 2000; 15: 111-31.

[11]. Johnston J. Econometric methods, 3rd ed. McGraw-Hill, New York 1984.

[12]. Kennedy P. Sinning in the basement: what are the rules? The ten commandments of applied econometrics. J. Econ. Surveys 16: 569-589.

[13]. Chatfield C. The initial examination of data. Journal of the Royal Statistical Soc. 1985; A148, 214-31.

[14]. World Bank. Sub- Saharan Africa: From crisis to sustainable growth. Washington, D.C. 1989a.

[15]. World Bank Policy Report. Adjustment in Africa: reforms, results and the road ahead. Oxford University Press 1994.

APPENDIX

Data Problems in Third World Countries

Applied econometric research is severely undermined by the unique nature of data problems in Third world countries. First and foremost, adequate and complete data are hard to come by. The paucity coupled with reliability of data are major challenges facing empirical work in less developed countries in general and African countries in particular [14]. Many studies have had to confront not only the lack of data but also the gross deficiencies of available statistics. In fact, data from corrupt regimes due to bad governance (failed states) are often misinforming and disheartening[3].

Until recently, most statistical institutions were undeveloped, lacking financial and manpower resources to collect and process the necessary data. Most statistics of developing countries are limited and sketchy on a wide range of variables [15]. For example, the scope of Africa's data before 1970 is not only limited but fragmented in various publications, articles and reports. This has restrained data analysis and narrowed the scope of applied econometrics.

[12] Note that low R^2 should not necessarily be a disappointment because they are frequently encountered in cross-sectional data with a large number of observations. Besides, a model with low R^2 may be quite acceptable based on other criteria (i.e., statistical significance of estimated coefficients, theoretical adequacy of the model, etc.).
[13] This is done by conducting unit root tests. Details of these tests are discussed in "Time Series Econometrics".

Most time series data for third world countries are short, inadequate and are mainly compiled on annual basis. Lower frequencies of data (i.e., semi-annual, quarterly, monthly, etc.) are limited. Data collection methods in many countries are still rudimentary. Modern facilities and technology for data analysis are rare and scarce [2]. The collection of information and data is undermined by low respondents' responses, high levels of illiteracy, poor logistics and general population malaise in freely providing required information especially where conflicting interests are involved [3]. The data collected or compiled are in most cases never synchronized for uniformity.

Though the range of data collection and compilation has considerably improved for many countries in the past twenty years, major problems still exist. Financial constraints in collecting and compiling especially primary cross-sectional and panel data are formidable. Reconciling information from various sources remains a haphazard task [2]. There are several inconsistencies in data arising from independent reporting often because of the time lag involved when countries make submissions to differing sources. Cases of incomplete or unreported data are common, arising mainly from failures of various bodies (i.e., countries, agencies, individuals, etc.) to cooperate. Complete and timely data are not readily available because of the disparity in lags taken to compile and report information.

While several proposals to handle Third world data problems for econometric research have been made, it should be recognized that Africa still needs a strong statistical data base in order to appreciate and tackle a myriad of its development constraints [14]. Hence, it is imperative that these countries strive to initiate and strengthen commitment for the development of statistical institutions in the region.

Author

Ocaya Bruno (bocaya@isae.mak.ac.ug) is the Chair, Department of Statistics and Actuarial Science, COBAMS, Makerere University, Kampala. He holds a PhD in economics, majoring in econometrics from Northwestern University, USA. He has a diverse and long teaching carrier in various fields of econometrics and mathematical economics. Dr. Ocaya has published a book entitled: "Basic Mathematics and Statistics for Econometricians". He also has two additional manuscripts ready for publication namely, (1) "Simplified Practical Econometrics" and (2) "Panel Data Econometrics".