

MODELING WITH TIME SERIES: ISSUES AND COMMON ERRORS

By

Kilishi A. Abdulkakeem

Department of Economics
University of Ilorin
e-mail: meetkilishi@yahoo.com
Phone: 08050221980

This version: March 2022

Department of Economics, University of Ilorin, Working paper series, No 07

Abstract

Time series data are widely used in empirical research but it is observed that most often the modeling is done with errors. It is common for researchers to use a given estimation method without considering the stochastic properties of the variable. When time series data are used in estimation without addressing the problem of stochastic innovations, the result may be biased with invalid inferential statistics. Hence, hypothesis test will be unreliable and conclusion misleading. This paper provides simple discussion of potential problems that could arise while modeling with time series. Practical step by step approaches of modeling time series variables at different circumstances is also discussed in the paper. It is concluded that necessary pre estimation tests should always be carried out before choosing the appropriate method.

Key words: Time Series, Time Trend, Nonstationary, Modeling

JEL Classification: C32, C51, C18, C82

This paper was presented at the 56th Annual Conference of the Nigerian Economic Society (NES), Sheraton Hotel & Towers, Abuja, 11th to 14th October, 2015

1. Introduction

It is a common task in the fields of economics, finance, geography, business, even political science and other related disciplines to develop models so as to uncover the dynamic paths in order to estimate, interpret, forecast and test hypothesis using time series variables. This led to huge branch of knowledge – time series analysis. The marriage between the statisticians and econometricians approaches of time series analysis gave birth to time series econometrics. Time series econometrics has almost become synonymous with empirical macroeconomics and financial econometrics (Greene, 2008). As noted by Greene (2008) the literature in this field has grown at remarkable rate more than any other branch of econometrics. Though, it is impossible to provide a comprehensive list of textbooks in the field, leading texts include Kirchgassner and Wolters (2007), Chatfield (2004) Hatanaka (2003), Harvey (1990) and for advanced treatment Enders (2010) and Hamilton (1994) among others. In addition, all good available econometrics text would often dedicate at least one chapter to discuss time series modeling. For instance, Wooldrige (2006) used four chapters of his book (Introductory Econometric: Modern Approach) to discussed regression analysis using time series data, while Gujarati and Porter (2009), Stock and Watson (2007), Hill, Griffiths and Lim (2012), and Greene (2008) dedicated two, three and four chapters respectively to discuss time series analysis.

The nature of time series data is such that some of the assumptions of classical regression model are most often violated. These violations and how to deal with them are discussed extensively in the literature. However, the notations and theoretical derivations in time series literature usually appear knotty and complicated making it difficult for applied analysts of time series data to follow. The aim of this paper therefore is to provide simple discussion without compromising substance, the common regression estimation problems that typically arise when modeling with time series data. Specifically, the paper (i) reviews the characteristics of time series variables that make modeling with time series using classical econometrics unreliable; (ii) examines the potential problems of time series regression; (iii) discusses the common errors researchers make while trying to address the time series regression problems; and finally (iv) elucidates the practical procedures to addressing problems associated with time series regression.

The remainder of the paper is arranged in six sections. Nature of time series variables in section two, review of Gauss Markov theorem and assumption in three, while issues with time series regression, practical procedure of modeling with time series, common errors in modeling with time series and concluding remarks are presented in sections four through seven respectively.

2. Nature of Time Series Variables

The data generation process of time series variables such as inflation, exchange rate, stock market prices, income growth, e.t.c. is such that there is observation dependency across time. Particularly it is true that past values of a series affect current value and current value affects future values, but current does not affect the past. It follows therefore that the best way to predict the future values of time series variables is to use its own history. This naturally leads to Autoregression (AR) model, a regression model that relates a time series variable to its past values. Using theory, empirical evidence and/or common sense, researchers usually add the contemporaneous and lag values of other time series variables to predict the variable of interest. The addition of lags of predictor variables leads to Distributive Lag (DL) model. Thus, inclusion of lags of both the dependent and the independent variables yields Autoregressive Distributed Lag (ARDL) model. The implications of ARDL model within the context of classical regression model are discussed in the next section.

The second and of course most important characteristics of time series variables which pose serious challenges to time series regression is time trend. A variable is said to be trending if its value is either increasing (upward trend) or decreasing (downward trend) over time. Trend can be deterministic or stochastic. Deterministic trend is when the values of a series change at a fixed rate over time while stochastic trend is when the rate of change in the values of a series over time varies (i.e. the rate of change is random).

A series that has time trend would usually exhibit nonstationarity. A variable is said to be nonstationary if the mean, or variance or both vary with time. A trending variable follows a random walk, that is, the current value of the variable is equal to its past value(s) plus a random component. If a variable follows a random walk, its variance changes over time, hence, the variable is not stationary. Therefore, a variable that is trending (particularly stochastic trend)

would naturally not be stationary. Both trending in itself and the product of trending – nonstationary – constitute some regression problems as discussed in section four of this paper.

3. Review of Gauss Markov Theorem and Assumptions

The Gauss Markov theorem states that:

- i. the Ordinary Least Squares (OLS) estimator of sample beta $\hat{\beta}$ (regression parameters) is an unbiased estimator of the true population beta β whether conditional or not on the independent variable(s);
- ii. the conditional variance $\hat{\sigma}^2$ is an unbiased estimator;
- iii. the condition variance $\hat{\sigma}^2$ is the best because it gives the smallest variance among other linear unbiased estimators; and
- iv. OLS estimator $\hat{\beta}$ is a consistent estimator (this is the large sample property, i.e. the asymptotic property).

The first three theorems are the finite sample properties while the fourth is the asymptotic (large sample) property. The finite sample properties hold under the following assumptions:

- 1) The stochastic process Y_t (dependent variable) follows a linear function (linear in parameters);
- 2) The mean of error term conditional on the independent variables is zero;
- 3) Absence of perfect collinearity between two different independent variables;
- 4) The conditional variance of the error term is the same for all time t ;
- 5) The conditional errors in two different time periods are uncorrelated; and
- 6) Given assumption 1 through 5, the OLS estimators follow normal distribution, and the standard error, t and F statistics are valid.

Extension of these assumptions for the theorem to hold in case of time series regression are:

- i. The variables are stationary
- ii. The independent variables are exogenous
- iii. The model is dynamically complete

Discussion of assumptions 1 to 5 is deliberately avoided in this paper, interested reader who requires further explanation should contact any of the textbooks mentioned in section one of this paper. It is however important to mention here that assumptions 1 to 3 are necessary for beta to be unbiased, while assumptions 4 and 5 plus i. are required for variance to be unbiased and be the best.

The additional assumptions (i. through iii.) are extensions of traditional classical assumptions that are required for the theorem (or even the assumptions) to hold when modeling with time series data. For instance, assumption 2 will fail if assumption iii is violated. In other words, if a model is dynamically incomplete, the assumption that the conditional mean of error term is zero will fail. A model is said to be dynamically complete if no lags of either the dependent or independent variables that is important in predicting the current/future values of the dependent variable is omitted. Similarly, if all independent variables are not strictly exogenous, the assumption of zero conditional mean will not hold. The stationary assumption is the time series version of Independent Identically Distributed (iid) assumption. Thus, if the stationary assumption is violated a number of the Gauss-Markov assumptions including the normality assumptions would not hold. The specific potential regression problems associated with time series regression are discuss in the following sections.

4. Issues with Time Series Regression

The three characteristics of time series variables (observational dependency, time trend and nonstationarity) discussed in section two are interconnected and the consequences are equally similar. However, specific problems associated with each characteristics are discussed in turn.

If the values of a series are correlated across different time periods, the assumption of absence of serial correlation (values of error term not related across time, assumption 5) would likely be violated. If this the only problem, OLS estimator would still be unbiased but variance is no longer the best. However, observational dependency naturally leads to dynamic ARDL model, then strict exogeneity assumption breaks down. In this case Gauss – Markov assumption 2 and 5 are jointly violated. Hence, the estimator of beta is biased and variance is no longer the best, (that is, not smallest).

Time trend in itself does not violate any of the Gauss – Markov assumptions. However, if two series are trending whether in same or opposite direction, regression result may show that they are strongly related when they are not actually related. What appears in the regression result is a trend relationship and not an actual relationship, this problem is generally known as spurious regression problem. But if the two series contain common stochastic trend, the series will have a long run relationship (cointegrated). This implies that the linear combination of two nonstationary series may be cointegrated. Meanwhile, time trend leads to the problem of nonstationarity which constitute more serious consequences as discussed in what follows below.

If series are not stationary, the assumptions of constant variance across time is violated. In addition, the estimator follows a non-normal distribution. Hence, the OLS estimator is still unbiased but will not longer be best and consistent. Therefore, hypothesis testing using standard error, t and F statistics as well as conventional confidence intervals would not be valid.

In summary, the principal potential problem pose by trending and nonstationary series is a large variance. Thus, forecast which is one of the main purpose of time series modeling, will be biased and inefficient, and conventional statistical inferences would be unreliable and misleading. Trending and nonstationarity do not affect the unbiased property of OLS estimator. OLS estimator becomes biased only when a model is dynamically incomplete.

5. Practical Procedure of Modeling with Time Series

Based on the preceding discussions, possible scenarios that researchers can encounter when modeling with time series can be grouped into four. First, the series have no time trend and are stationary (this is a very rare case), second, series are trending but stationary (also rare), third, series are trending, not stationary and not cointegrated, and fourth, series are trending and not stationary but are cointegrated.

It is important to briefly discuss the concept of differencing because this has direct bearing with practical remedies to time series regression problems. One fundamental way to deal with trending and nonstationary is by differencing. Differencing is finding the change in the value of a variable between different time periods. The first difference for instance of a variable say Y_t is the change in the value of Y between time $t - 1$ and time t , that is, $\Delta Y_t = Y_t - Y_{t-1}$. Consider a simple model of a variable that has stochastic trend such that the current value of the variable is

equal to its past value plus random term. That is: $Y_t = Y_{t-1} + u_t$, where u_t is random term. This means that the value of the series today is its value yesterday plus an unpredictable change. The path followed by the series consist of random steps u_t , then this path is called 'random walk'. If a series follows random walk such as Y_t , the variance increases with time making the distribution of Y_t to vary over time, hence Y_t is not stationary. However, the first difference of pure random walk is stationary. The number of times a series is differenced before it becomes stationary defines the order of integration. If a series becomes stationary after been differenced once, its said to be integrated at order one, professionally written as I(1) and if it is differenced twice is I(2) and so on. A series that is stationary at level is a I(0) series.

The practical procedure of modeling time series is summarised schematically in figure 1 and table 1 below. The first step is to test the presence of trend and unit root (unit root is synonymous to nonstationarity), or the order of integration. If series are stationary at level without time trend, all Gauss-Markov assumptions will be intact and OLS will produce reliable result with valid statistics for hypothesis testing. However, if series are stationary but have trend, then OLS should be estimated with de-trending. In this case, is either to estimate regression with time variable added as one of the regressors or de-trend the series first before running regression with de-trended series.

If series are not stationary, particularly if they are I(1), researcher needs to proceed to Cointegration test. It should be mentioned here that the available techniques for Cointegration test in E-views software (Engle-Granger, Phillips-Ouliaris, and Johansen) assume series are I(1), while bound testing procedure is capable of combining I(0) and I(1) series. Engle – Granger and Phillips – Ouliaris Cointegration tests are suitable for single equation models, while Johansen Cointegration test is for system of equations model. A model is referred to as system of equations model if both the dependent and independent variables influence each other and a model is single equation when only the independent variables affect the dependent but the dependent variable does not in turn affect the independent variable(s).

If nonstationary series [I(1) series] are not cointegrated, the series have to be modeled as first difference ARDL model if it is a single equation model, or first difference VAR (Vector Autoregression) if is a system of equations model. Since, series are I(1), first differencing is necessary to make the series stationary series (difference stationary) and inclusion of the lags

will account for the impact of the history of the series. However, if it is the combination of I(0) and I(1) series, when there is no Cointegration, the I(0) series will enter the model at their levels without differencing.

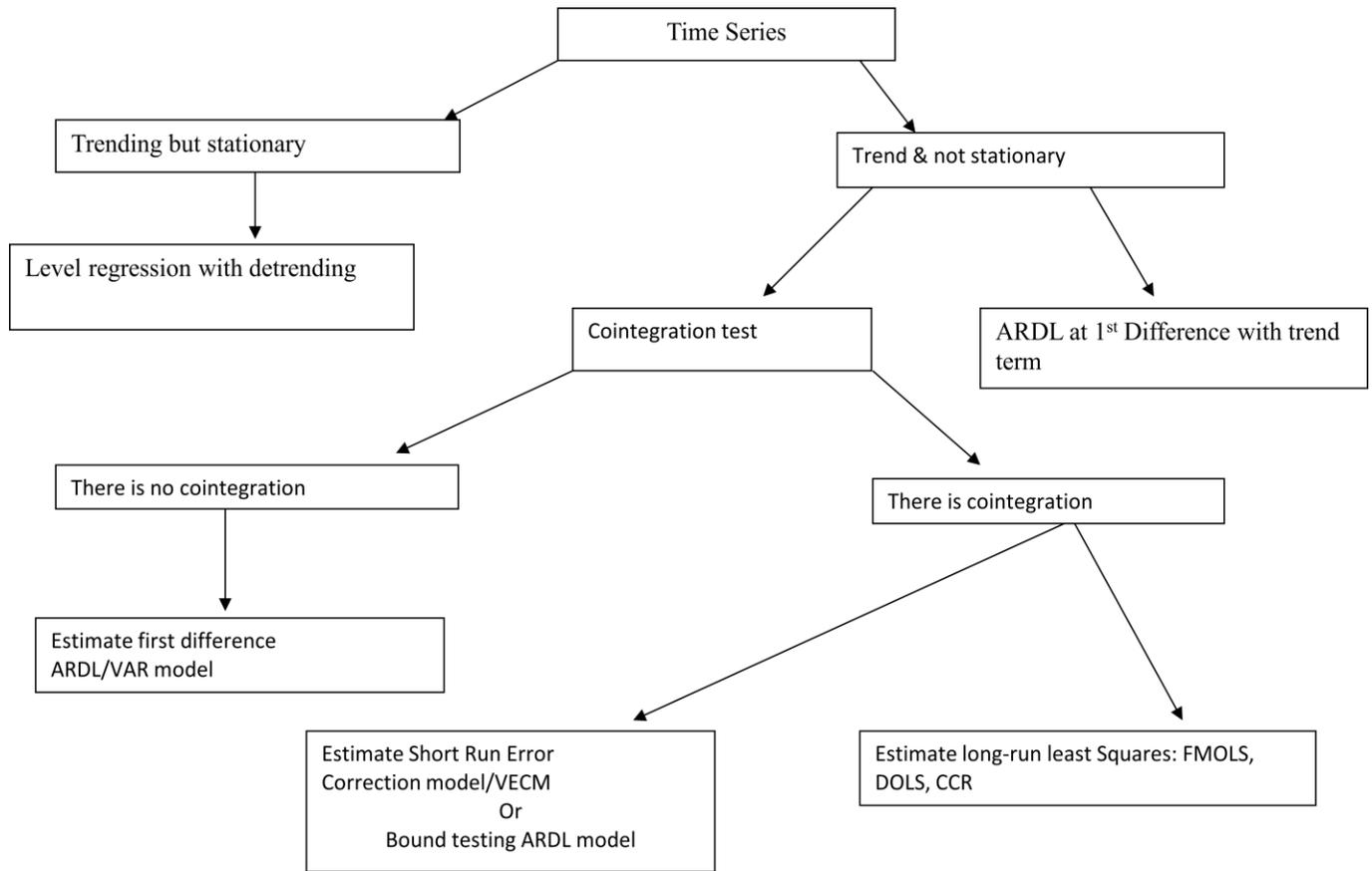
On the other hand, if series are cointegrated, there are two options if it is a single equation model. First, is to estimate short run Error Correction Model (ECM). Second, is to estimate long run cointegrating regression. The ECM is first difference ARDL model with the inclusion of error correction term which account for the long-run characteristics of the series. Hence, series enter ECM mode as difference stationary series. The second option consists of three fully efficient estimation techniques for long- run cointegrating regression, these are Fully Modified OLS (FMOLS), Dynamic OLS (DOLS) and Canonical Cointegrating Regression (CCR). FMOLS employs a semi-parametric correction to eliminate the problems caused by the long run correlation between the cointegrating equation and stochastic trend. CCR is closely related to FMOLS, but instead employs stationary transformations of the series so as to obtain least squares estimates that will eliminate the long run dependence between the cointegrating equation and stochastic trends. While DOLS involves augmenting the cointegrating regression with lags and leads of the independent variables so that the resulting cointegrating equation error term is orthogonal to the entire history of the stochastic trends.

The bound testing ARDL modeling is suitable for single equation model when series are combination of I(1) and I(0). It is a short – run error correction technique that combines the short – run and long – run characteristics of series. On the other hand, 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.

Table 1: Characteristics of Series and Appropriate Methods

Cases	Time trend	Non-stationary	Cointegration	System of Equations	Method
Case 1	No	No	No	No	OLS at Levels
Case 2	Yes	No	No	No	OLS with Detrending
Case 3	Yes	Yes	No	No	First Difference ARDL
Case 4	Yes	Yes	Yes	No	Short-run ECM, Long-run OLS
Case 5	No	No	No	Yes	VAR
Case 6	Yes	Yes	Yes	Yes	VECM

Figure 1: Practical Procedures



6. Common Errors

It is obvious that empirical analysts of time series are aware of the potential problems that may arise as result of the presence of stochastic trend in variables. Almost everyone who uses time series would have conducted unit root test to ascertain the order of integration of series. However, majority fall into one or more errors when modeling series with stochastic trend. The discussion in this section is based on the errors observed in a number of post graduate students research papers and articles reviewed by the author for some national journals.

Generally, the observed errors can be grouped into five categories, (i) model specification errors; (ii) deciding the modeling procedure before or without pre – test; (iii) use of inappropriate

techniques; (iv) disconnect between specified and estimated models; and (v) wrong solutions to genuine concerns.

Specification errors can take different forms such as omission of relevant variables, inclusion of irrelevant variables, wrong functional forms, and measurement errors. The one most relevant in this discussion is dynamic incomplete models which falls under omission of relevant variables. A model is said to be dynamically incomplete if lags of either the dependent or independent variables that are useful in explaining current and future values of the dependent variable are omitted. It is therefore important while modeling any version of ARDL models to conduct lag length test so as to ensure that dynamically complete model is obtained. This is even more important if the purpose of the model is forecasting. Majority of people only choose the number of lags arbitrarily which may lead to omission of useful lags or inclusion of irrelevant lags.

The second and third categories of errors are related; Some people would have decided the method to use right from the topic, hence go on to use it irrespective of the characteristics of variables. It is useful at this point to mention some of the wrong use of techniques observed in some empirical papers. First, people freely use Johansen Cointegration test even if model is single equation. As mentioned earlier Johansen is suitable for system of equations model, using it for single equation will amount to error. Second, including series of different order of integration into cointegration test. Sometimes, $I(2)$ series are also included. None of the available tests in the software can handle $I(2)$, not even bound testing procedure. Bound testing is only useful when series are combination of $I(0)$ and $I(1)$. Third, modeling $I(1)$ without differencing, this is not appropriate because one or more problems can arise if regression is estimated with nonstationary series. Sometimes people include $I(2)$ variable by differencing it twice, though, there is no theoretical position against this but the long – run information of such variable may be lost completely. Even interpreting coefficients of such variables is difficult.

Most often there is disconnect between specified and estimated models. Researchers may start with single equation specification but end up estimating VAR and VECM model. In some cases, the final restricted error correction model estimated is different from the unrestricted model used for the bound testing Cointegration. The model used for the bound test is supposed to be the best model that is dynamically complete, estimating any model for the same data will be wrong.

Some researchers would identify real problem but often use inappropriate method to solve it. For example, a researcher once raise concern about endogeneity of the independent variables which is a genuine concern because if the independent variables are endogenous, regression estimate will produce biased result. The researcher in question attempted to solve this by using DOLS (Dynamic OLS). Though, the concern is genuine but the solution is wrong because DOLS does not have the capacity to solve endogeneity problem. Techniques such as Instrumental Variable (IV), Two Stage Least Squares (2SLS), Generalized Method of Moment (GMM) e.t.c. are appropriate techniques when there is endogeneity problem or estimate system of equations model.

7. Concluding Remarks

Modeling with time series data is popular in empirical research in the fields of economics, finance, business, and other related discipline. However, the modeling is most often done erroneously without following appropriate procedure or applying the appropriate techniques. Some of the common mistakes people make and the practical approaches to modeling time series variable are discussed in this paper. It is important that when modeling with time series to conduct pre-estimation tests so as to confirm the characteristics of the series which will indicate the appropriate technique to use. It is also ethical to conduct post estimation tests in order to ensure that the problem the estimation technique is supposed to solve is actually solved.

This paper tried to summarized the major practical approaches of modeling time series variable. It is however, not possible to exhaust all the issues concerning time series modeling in a single paper. Moreover, the paper only mentioned appropriate methods given different characteristics of time series, but how to practically implement these estimation method is not discussed. For more details, any of the textbooks mentioned in section one can be consulted. The E-Views user guide is also a pretty good material for extensive details, particularly in how to implement the methods discussed using E-Views.

References

- Chatfield C. (2004), *The Analysis of Time Series: An Introduction*, Florida: Chapman and Hall/CRC,
- Enders W. (2010), *Applied Econometrics Time Series*, New York : John Wiley and Sons,
- 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 University Press
- Hatanaka M. (2003), *Time Series Based Econometrics*, New York: Oxford University Press
- Harvey A. (1990), *The Econometrics Analysis of Time Series*, Cambridge: MIT Press
- Hill C. R., Griffiths E. W. and G. C. Lim, (2012), *Principles of Econometrics*, Hoboken: John Wiley & Sons, Inc
- Kirchgassner G. and J. Wolters (2007), *Introduction to Modern Time Series Analysis*, New York: Springer Berlin Heidelberg,
- Stock H. J. and M. W. Watson, (2007), *Introduction to Econometrics*, Boston: Pearson Education, Inc.
- Wooldridge M. J., (2006), *Introductory Econometrics: A Modern Approach*, Toronto: Thomson South-Western