A Simple Integration Order Test: An Alternative to Unit Root Testing

C. G. Amaefula

Abstract — The paper introduces order of integration test (OIT) which serves as a simple alternative to unit root test built generally using auxiliary autoregressive AAR(3) model. The parametric boundary conditions necessary and sufficient for testing the null hypothesis that the non-stationary variable under test is integrated order zero I(0) were estimated via generalized least squares (GLS). The decision on the hypothesis is evaluated using t-statistic. The test procedure was applied to a simulated non-stationary series \( y_t \) of sample size \( n = 2000 \) and a known non-stationary time series data \( y_t \) with two unit roots. The results showed that \( y_1 \) is integrated order one \( I(1) \) and \( y_2 \) is \( I(2) \). These results were confirmed by Augmented Dickey Fuller (ADF); Phillips-Perron (PP); Kwiatkowski, Phillips, Schmidt, and Shin (KPSS); Elliot, Rothenberg, and Stock Point Optimal (ERS) and Ng and Perron (NP) unit root tests. For logarithm transformed variable, the divergent opinions of other unit root tests in clear-cut solution of the integrated order of such variable makes the new test procedure a better alternative. Nevertheless, the simplicity and aptness of the integration order test give it leverage over conventional methods of unit root test.

Key words — AAR(3), GLS, Integration order test, unit roots, parametric boundary conditions.

I. INTRODUCTION

The order of integration of a series is very important in Econometrics and Time series modelling. Researches on unit root testing have attracted attention for the past three decades. Detecting the order of integration of a data series becomes somewhat difficult except a unit root test is carried out. If a data series is non-stationary, and a unit root test has confirmed that the level series of such data is not stationary then, after the first differencing and the data series becomes stationary, such data series is said to be integrated order one \( I(1) \). But if the second differencing is needed to achieve stationarity, such data series is said to be integrated order two \( I(2) \). This means that not less than a double unit root tests is needed before confirming that a non-stationary series is \( I(1) \) or \( I(2) \).

Generally, the integration order of a variable is denoted by \( I(d) \), where \( d \) is the number of differencing needed to make the non-stationary variable stationary. If a variable is stationary, then, it is said to be integrated order zero \( I(0) \). The series could be non-stationary because of random walk, drift, or trend. A series is said to be stationary if its mean and autocovariances do not depend on time. Any series that is not stationary is said to be non-stationary. According to [5], Regressing a non-stationary variable on a deterministic trend generally does not yield a stationary variable (instead the series needs to be differenced prior to processing). Thus, using standard regression techniques with non-stationary data can lead to the problem of spurious regressions involving invalid inferences based on t- and F-tests. A non-stationary variable becomes stationary after it is differenced (although not necessarily just by first-differencing — it will be shown that the number of times a variable needs to be differenced in order to induce stationarity depends on the number of units roots it contains). According to [9], differencing will handle the transformations to stationarity. The question of whether a variable is stationary depends on whether it has a unit root. The Augmented Dickey-Fuller (ADF) test build a parametric correction for higher-order correlation, and it assumed that the \( y \) series follows an AR(p) process and adding \( p \) lagged difference terms of the dependent variable \( y \) to the right-hand side of the test regression. Moreover, while the assumption that \( y \) follows an autoregressive (AR) process may seem restrictive, [12] showed that the (ADF) test is asymptotically valid in the presence of a moving average (MA) component, provided that sufficient lagged difference terms are included in the test regression. [4] propose a simple modification of the (ADF) tests in which the data are detrended so that explanatory variables are “taken out” of the data prior to running the test regression. [11] propose an alternative (nonparametric) method of controlling for serial correlation when testing for a unit root. The PP method estimates the non-augmented DF test equation and modifies the t-ratio of the \( \alpha \) coefficient so that serial correlation does not affect the asymptotic distribution of the test statistic. According to [13], Philips type of test has poor size properties (i.e., the tendency to over reject null hypothesis when it is true) when the underlying data generating process has negative moving average (MA) components and MA terms are present in many macroeconomic time series.

The [6] introduced a test that differs from the other unit root tests called the KPSS test, which has a null of stationarity of a series around either mean or a linear trend; and the alternative assumes that a series is non-stationary due to presence of a unit root. In this respect it is innovative in comparison with earlier Dickey-Fuller test, or Perron type tests, in which null hypothesis assumes presence of a unit root. [10] construct four test statistics that are based upon the GLS detrended data \( y_t^* \). These test statistics are modified forms of Phillips and Perron \( Z_u \) and \( Z_t \) statistics, the [1] \( R \) statistic, and the ERS Point Optimal statistic.

Published on July 24, 2021.
C. G. Amaefula, Federal University Otuoke, Nigeria. (corresponding e-mail: wordwithflame@gmail.com)

DOI: http://dx.doi.org/10.24018/ejmath.2021.2.3.22

Vol 2 | Issue 3 | July 2021
The study presents a simple analytical method of testing for integration order using AAR(3) process with optional intercept term. The remaining part of the paper is arranged as follows: section II presents the materials and methods, section III presents data analysis and results and section IV deals with summary and conclusion.

II. MATERIALS AND METHODS

The method in this study is evidence-based approach relies on direct observation and experimentation in the acquisition of new knowledge. However, this section presents the methodology of the proposed order of integration test mechanism that can be used as a simple alternative to unit root test.

A. A Non-stationary Series

A series is said to be non-stationary if it is integrated order one I(1) or order two I(2). The order of integration is the number of unit roots contained in the series, or the number of differencing needed to make the series stationary. Consider a non-stationary autoregressive (1) process of the form:

$$y_t = \rho y_{t-1} + \epsilon_t$$  \hspace{1cm} (1)

where $\rho = 1$, (1) becomes a random walk. $\epsilon_t$ is a stationary random disturbance term, since:

$$\begin{align*}
y_t - y_{t-1} &= \epsilon_t \\
(1-L)y_t &= \epsilon_t \\
\nabla y_t &= \epsilon_t
\end{align*}$$  \hspace{1cm} (2)

The first difference of $y$ denoted by $\nabla y_t$ at time $t$ in (2) is stationary, that is I(0) and $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$. Sometimes, unit root might still be present in a differenced series $\nabla y_t$, this means the variable $y$ is integrated order two I(2), hence, second differencing is required to make it stationary. That is:

$$\begin{align*}
\nabla y_t &= \rho^* \nabla y_{t-1} + \epsilon_{t1} \\
\nabla^2 y_t - \nabla y_{t-1} &= \epsilon_{t1} \\
(1-L')y_t &= \epsilon_{t1} \\
\nabla^2 y_t &= \epsilon_{t1}
\end{align*}$$  \hspace{1cm} (3)

where $\rho^* = 1$ and $d = 2$. In (3), $y_t$ is said to be integrated order two I(2) and $\epsilon_{t1}$ is a stationary random disturbance term. The null hypothesis $H_0 : \rho = 1$ or $\rho^* = 1$ can be tested against the alternative $H_a : \rho < 1$ using the critical values of t-statistic. If the null hypothesis cannot be rejected, the data generating process is said to have unit root and to be nonstationary. Therefore, the two-sided significance test performed is for the statistical significance of $\rho - 1$ or $\rho^* - 1$ as the case may be. The test:

$$t = \frac{p - 1}{s(e/a)}$$  \hspace{1cm} (4)

The test resembles a $t$-test. The null hypothesis $H_0 : \rho = 1$ is rejected if $|t| > a_t$, where the value of $a_t$ depends on the sample size and which other parameters are in the equation.

According to [3] under the null hypothesis of a unit root, this statistic does not follow the conventional $t$-distribution, and they derived asymptotic results and simulated critical values for various test and sample sizes. Thereafter, [7], [8] used larger set of simulations than those tabulated by Dickey and Fuller.

B. Detecting Integration Order $I(d)$ – Differencing Frequency

One common way of detecting the number of times a non-stationary series $\{y_t\}$ can be differenced to become stationary most often is allowing $d = 1$, then test for unit root using any one of the several tests available in literature such as the Augmented Dickey-Fuller (ADF) test, Phillips-Perron (PP) test, Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test and so on. If after first difference, $\{y_t\}$ is still non-stationary, the second differencing ($d = 2$) is expected to render the underlying series I(0). However, it is conventional practice that $d$ cannot exceed 2. It has been aforementioned that the number of differencing needed to make a non-stationary series stationary is known as the integration order.

C. Detecting Integrated Order one I(1) – Presence of One Unit Root

However, the differencing frequency to render a non-stationary series I(0) can be visualized using auxiliary autoregressive (AAR) model. Consider a non-stationary series with lagged endogenous regression of order two with a constant term of the form:

$$y_t = \beta_0 + \varphi tr + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \epsilon_t$$  \hspace{1cm} (5)

where, $\varphi$ is the coefficient of the trend parameter $tr$, $\beta_0$ is the constant term (intercept) and it is optional to include it. The $\epsilon_t$ is the random error term. $\beta_1$ and $\beta_2$ are the autoregression coefficients. But if the level series $\{y_t\}$ is integrated order one I(1), then it is conditionally sufficient that the absolute value of $\beta_1$ must be greater than or equal to unity, that is, $|\beta_1| \geq 1$, $\beta_2$ must be strictly less than one, that is $|\beta_2| < 1$ and it is necessary that the ratio of $|\beta_1|/|\beta_2| > 1$. For a differenced $\{y_t\}$, $y_t - y_{t-1} = (1-L)y_t = \nabla y_t$, (5) is rewritten as:

$$\nabla y_t = \beta_0 + \varphi tr + \beta_1 \nabla y_{t-1} + \beta_2 \nabla y_{t-2} + \epsilon_t$$  \hspace{1cm} (6)

If the level series is integrated order one I(1) after the first difference, the absolute values of $\beta_1$ and $\beta_2$ in (6)
should be strictly less than unity, that is, $|\beta_1|<1$ and $|\beta_2|<1$, respectively. And the random disturbance term $e_t$ should be stationary.

**D. Detecting Integrated Order One I(2) – Presence of 2 Unit Roots**

Consider (4) again, if the level series $\{y_t\}$ is integrated order two I(2), then it is expected that $|\beta_1|>1$ and the value of $|\beta_2|\geq1$. Situation could arise where $|\beta_2|\equiv1$ or the value of $|\beta_2|$ is approaching unity. If such situation arises, increase the lag order of (5) to be of the form:

$$y_t = \beta_0 + \rho \tau + \beta_2 y_{t-2} + \beta_3 y_{t-3} + e_t \quad (7)$$

For an integrated order two I(2) series $\{y_t\}$, it is conditionally sufficient that $|\beta_1|>1$, $|\beta_2|\geq1$ and $|\beta_3|<1$. We can as well use (6) to ascertain whether a series $\{y_t\}$ is still non-stationary after first difference. Obtaining the absolute value of $\beta_2$ to be greater than or equal to unity, we say that $\{y_t\}$ is integrated order two I(2).

Generally, for clarity purpose, (7) should be used for order of integration test. The parametric boundary conditions for I(1) are as follows: $|\beta_1|\geq1$, $|\beta_2|<1$, $|\beta_3|<1$ and $|\beta_1|>1$ and for I(2) it is expected that $|\beta_1|>1$, $|\beta_2|\geq1$, $|\beta_3|>1$ and $|\beta_1|\geq1$. However, detecting the order of integration for non-stationary series is synonymous with detecting the number of unit roots present in the same series. If a non-stationary series is I(1), it means there is only one unit root in that series. Where a non-stationary series is I(2), it means the series has two unit roots.

**E. Proposition 1**

Consider an AAR(3) model as given in (7), if $\{y_t\}$ is I(1) and $\{\beta_2, \beta_3\} \in (-1, 1)$, then it suffice that $|\beta_2|>1$. 

**Proof:**

AAR(3) as presented in (7) can be re written as:

$$\left(1-\beta_1 L-\beta_2 L^2-\beta_3 L^3\right)y_t = \beta_0 + \rho \tau + e_t \quad (7.1)$$

Let $\Phi = \beta_0 + \rho \tau$ be the deterministic term so that our focus can be on the polynomial of order 3. Then (7.1) becomes:

$$\left(1-\beta_1 L-\beta_2 L^2-\beta_3 L^3\right)y_t = \Phi + e_t \quad (7.2)$$

Equating the polynomial to zero, we have:

$$1-\beta_1 L-\beta_2 L^2-\beta_3 L^3 = 0 \quad (7.3)$$

Equation (7.3) becomes:

$$\beta_1 L + \beta_2 L^2 + \beta_3 L^3 = 1 \quad (7.4)$$

For $\beta_1 = 0$ and $\beta_3 = 0$ in (7.4),

$$\beta_1 L = 1 \quad (7.5)$$

From (7.5), it implies that $|\beta_1| \geq 1$ if $L \in [-1, 1]$. So, assuming that $-1 \leq L \leq 1$ then,

$$|\beta_1| = \begin{cases} 1 & \text{if } L = \pm 1 \\ 0 & \text{if } L = 0 \\ >1 & \text{otherwise} \end{cases} \quad (7.6)$$

Note that $L$ is the lag polynomial at order 1 and cannot be zero. Hence, $|\beta_1| \geq 1$ and without loss of generality, $|\beta_1| \in [1, \infty)$.

**Corollary 1:** For an I(1) series, given that $|\beta_1| \geq 1$ and $\{\beta_2, \beta_3\} \in (-1, 1)$, then $\left|\frac{\beta_1}{\beta_2}\right| > 1$.

**Proof:**

For any $\varepsilon \in (0, 1)$, given that $|\beta_1| \geq 1$ then,

$$|\beta_2 - \varepsilon| < \frac{|\beta_2 - \varepsilon|}{|\beta_2 - \varepsilon|^2} < \frac{|\beta_2 + \varepsilon|}{|\beta_2 + \varepsilon|^2} < \frac{|\beta_2 + \varepsilon|^2}{|\beta_2 + \varepsilon|^2} < |\beta_2| \quad (7.7)$$

Given that $\beta_2, \beta_3 \in (-1, 1)$, then $|\beta_2| < |\beta_3|$, it suffice to say that,

$$1 \leq |\varepsilon| < \left|\frac{\beta_1}{\beta_2}\right| \quad (7.8)$$

If (7.8) is true, then $\left|\frac{\beta_1}{\beta_2}\right| > 1$.

**F. Proposition 2**

Consider an AAR(3) model as given in (7), if $\{y_t\}$ is I(2) and $|\beta_3| \in (0, 1)$ then it is sufficient that, $1 \leq |\beta_2| < |\beta_3|$.

**Proof:**

Consider (7.4), it can be shown that:

$$\beta_2 = \frac{1-\beta_1 L-\beta_3 L^3}{L^2} \Rightarrow |\beta_2| = \left|1-\beta_1 L-\beta_3 L^3\right| \quad (7.7)$$

Since $\beta_2$ is the coefficient of $L^2$ (which invariably is the notation for lag 2). It is not confusing to let $L = \pm 2$ and given that $|\beta_1| > 1$. Then, at any admissible value of $\beta_2 \in (0, 1)$ it can be shown that:

**DOA:** http://dx.doi.org/10.24018/ejmath.2021.2.3.22
\[ \frac{1 - \beta_1 L - \beta_2 L^2}{L^2} > 0 \]

Hence, \( |\beta_2| \geq 1 \).

**Corollary 2:** For an I(2) series, if \( 1 \leq |\beta_1| < |\beta_2| \) and \( |\beta_2| \in (0, 1) \), \( \beta_1 \in (0, 1) \), and \( \epsilon \in (0, \frac{1}{2}) \), it is necessary that \( |\beta_2| > 1 \).

**Proof:** Given that \( 1 \leq |\beta_1| < |\beta_2| \) and \( \beta_2 \in (0, 1) \), for any \( \epsilon \in (0, \frac{1}{2}) \), then,

\[
\left| \frac{\beta_2}{\beta_1} \right| \leq \left| \frac{\beta_1 - \epsilon}{\beta_2} \right| < \left| \frac{\beta_1 - \epsilon}{\beta_1 - \epsilon} \right| < \left| \frac{\beta_1}{\beta_1} \right|
\]

(7.8)

Since \( \frac{|\beta_2| - \epsilon}{|\beta_1| - \epsilon} > 1 \), it can be verified therefore, that \( \frac{|\beta_2|}{|\beta_1|} > e^\epsilon \).

**Corollary 3:** For an I(2) series, if \( 1 \leq |\beta_2| < |\beta_1| \) and \( \beta_2 \in (0, 1) \), then \( |\beta_2^2| - |\beta_1^2| > 0 \).

**Proof:** Given that \( 1 \leq |\beta_2| < |\beta_1| \) and \( \beta_2 \in (0, 1) \), for any \( \epsilon \in (0, \frac{1}{2}) \), then it is sufficient to show that the

\[
\frac{|\beta_2^2 - \beta_1^2 + \epsilon^2|}{|\beta_2^2 - \beta_1^2 + \epsilon^2|} > 1
\]

That is,

\[
|\beta_2^2 - \beta_1^2 + \epsilon^2| > |\beta_2^2 - \beta_1^2 + \epsilon^2|
\]

(7.9)

However, since \( |\beta_2^2 - \beta_1^2 + \epsilon^2| > |\beta_2^2 - \beta_1^2 + \epsilon^2| \), it suffices to say that \( \frac{|\beta_2^2 - \beta_1^2 + \epsilon^2|}{|\beta_2^2 - \beta_1^2 + \epsilon^2|} > e^{\epsilon^2} \).

Note that **corollary 3** can be proved using triangular inequalities as follows.

Since \( |\beta_2^2 - \beta_1^2| \geq |\beta_1 - \beta_2| \), it is evident to show that \( |\beta_2^2 - \beta_1^2| > 0 \). So based on triangular inequalities \( |\beta_2^2 - \beta_1^2| \leq |\beta_2^2 - \beta_1^2| + |\beta_2^2 - \beta_1^2| \).

Taking the probability and expectation of both sides of the inequality and for any \( \epsilon > 0 \),

\[
p\left(E\left[\beta_2^2 - \beta_1^2 > \epsilon\right]\right) \leq p\left(E\left[\beta_2^2 - \beta_1^2 > \frac{1}{\epsilon}\right]\right) + p\left(E\left[\beta_2^2 - \beta_1^2 > \frac{1}{\epsilon}\right]\right) \to 1
\]

And thus, \( \lim_{\epsilon \to 0} p\left(E\left[\beta_2^2 - \beta_1^2 > \epsilon\right]\right) = 1 \).

Alternative way of proving proposition 2 is by differencing the non-stationary series \( \{y_t\} \) under study once so that (7.1) can be rewritten as:

\[
(1 - \beta_1 L - \beta_2 L^2 - \beta_3 L^3)\n\}

(7.10)

Then simply repeat proposition 1 and Corollary 1, respectively.

**G. Order of Integration Test Hypotheses**

In order to test the null hypothesis that the integration order of the variable \( y \) is equal to zero against the one-sided alternative that it is greater than or equal to unity. Using (6), we state the following hypotheses:

\[
H_{00} : \beta_2 < 1 \quad \text{(it is integrated order zero)}
\]

(7.11)

\[
H_{10} : \beta_2 > 1 \quad \text{(it is integrated order one)}
\]

(7.12)

\[
H_{01} : \beta_2 < 1 \quad \text{(it is integrated order zero)}
\]

(7.13)

\[
H_{11} : \beta_2 > 1 \quad \text{(it is integrated order one)}
\]

(7.14)

The test statistic for the above hypotheses is given as:

\[
t_j = \frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)}
\]

(8)

which has a Student’s t-distribution with \( n - p - 1 \) degrees of freedom. The test is carried out by comparing the observed value with the appropriate critical value \( t_{n-p-1, \alpha} \), which is obtained from the t-table. Where \( \alpha \) is the significance level. \( H_0 \) is to be rejected at the significance \( \alpha \) level if:

\[
|t_j| \geq t_{n-p-1, \alpha}
\]

(9)

where \( p\left(t_j\right) \) is the \textit{p-value} of the test, which is the probability that a random variable having a Student t-distribution, with \( n - p - 1 \) is greater than \( |t_j| \) (the absolute value of the observed value of the t-test). A criterion similar to that in (8) is to compare the p-value of the test with \( \alpha \) and reject \( H_0 \) if this probability value is less than the size of the test, say 0.05, we reject the null hypothesis.

**H. Method of Estimation Using Generalized Least Squares**

The method of estimation adopted is the generalized least squares method and it is presented briefly as follows:

(7) can be rewritten in matrix form as:

\[
y = Y'\beta + e
\]

(10)

where \( Y' \) is the vector of lagged dependent variables. The \( E(e) = 0 \) and \( \text{var-cov}(e) = \sigma^2V \). And where \( \sigma^2 \) is
known, V represents the assumed structure of variances and covariances among the random errors $e$. Under the given condition of the variance–covariance of the error terms $e$, $\beta$ 's can be estimates using:

$$\beta = (\mathbf{Y}' \mathbf{V}^{-1} \mathbf{Y}^{-1})^{-1} \mathbf{Y}' \mathbf{V}^{-1} \mathbf{y} \tag{11}$$

where var-cov($\beta$) = $\sigma^2 (\mathbf{Y}' \mathbf{V}^{-1} \mathbf{Y}^{-1})^{-1}$. If it is assumed that the variance $\sigma^2$ of each error term is the same over time and the error terms are mutually uncorrelated, then the V matrix reduces to identity matrix. And if there is heteroscedasticity as well as autocorrelation, then the V matrix will have entries on the main diagonal as well as on the off diagonal.

I. Unit Root Test

Unit root test is a conventional test for ascertaining the order of integration of any underlying time series data or unstructured data. In other words, unit root test analysis specifies whether any given data structure is stationary or nonstationary.

There are lots of unit root tests developed over time, for the literatures and technical details, see [1], [3], [4], [6], [10]-[13]. These unit root test sometimes are divergent and contradict each other on the order of integration of some data series, most especially logarithm transformed data series.

Explicit example where results of unit root test on a time series data as regards these unit root tests will be presented in section III.

III. DATA ANALYSIS AND RESULTS

The two data sets used in this study include simulated sample of 2000 observations and an empirical data on service sector output obtained from [2] statistical bulletin consisting of 39 observations spanning from 1981 to 2019. The empirical data on services sector output used is known to be integrated order two I(2).

A. Integration of Order One I(1) Analysis Using Simulated Data

The simulated data with drift is analysed for order of integration and the result is presented below:

$$\mathbf{y}_t = 0.3369 + 0.0006 + 1.3022 \mathbf{y}_{t-1} + 0.0135 \mathbf{y}_{t-2} + 0.3154 \mathbf{y}_{t-3} + \mathbf{e}_t \tag{12}$$

The result in (12) indicates that $\mathbf{y}_t$ is I(1) since $|\beta_1| = 1.3022 > 1$, $|\beta_2| = 0.0135 < 1$, $|\beta_3| = 0.3154 < 1$ and $|\beta_1| = 0.3369$ is significant under 5% level as expected. This result reveals that there is only one unit root present in the simulated series $\mathbf{y}_t$. Therefore, first differencing can be used to achieve stationarity of $\mathbf{y}_t$. It was stated earlier that $\beta_0$ (the constant term) is optional. Not including the constant term does not affect the integration order of the variable under study. The result without the constant term is presented in (13) below:

$$\mathbf{y}_t = 0.0019 + 1.4049 \mathbf{y}_{t-1} + 0.0140 \mathbf{y}_{t-2} - 0.4208 \mathbf{y}_{t-3} + \mathbf{e}_t \tag{13}$$

This result also confirms the integration order of $\mathbf{y}_t$. Giving that $|\beta_1| = 1.4049 > 1$, $|\beta_2| = 0.0140 < 1$, $|\beta_3| = -0.4208 < 1$ and $\frac{|\beta_1|}{|\beta_3|} = 3.3950 > 1$. And since the p-value (0.0000) of $\beta_1$ is significant under 5% level, the hypothesis that ($H_0$: $\beta < 1$) the variable is I(0) is rejected in favour of the alternative. Hence, the series $\mathbf{y}_t$ is I(1).

B. Integration of order two I(2) Analysis using Empirical Data.

The analysis of detecting I(2) or the presence of two unit roots in a set of data is demonstrated using service sector output data from CBN and the results is presented in (14) below:

$$\mathbf{y}_{t2} = -684.9050 + 69.4891 + 1.8847 \mathbf{y}_{t2-1} \tag{14}$$

The regression result in (14) reveals that the trend parameter value (69.4891) is significant under 5% level. The coefficients $|\beta_1| = 1.8847 > 1$, $|\beta_2| = -1.2293 > 1$, $|\beta_3| = 0.3586 < 1$, $\frac{|\beta_1|}{|\beta_2|} = 1.5331 > 1$ and $\frac{|\beta_2|}{|\beta_3|} = 3.4281 > 1$. The values of $\beta_1$ and $\beta_2$ are significant under 1% and 5% levels, respectively. This result entails that $\mathbf{y}_{t2}$ is I(2). Conventionally, $\mathbf{y}_{t2}$ can be stationary should the second differencing be applied. So, there are two unit roots present in service sector output data. Omitting the constant term in (14), the result becomes as giving in (15) below:

$$\mathbf{y}_{t2} = 27.6504 + 1.9532 \mathbf{y}_{t2-1} - 1.2736 \mathbf{y}_{t2-2} + 0.3436 \mathbf{y}_{t2-3} + \mathbf{e}_t \tag{15}$$

The choice to omit the $\beta_0$ as shown in (15) does not change the order of integration or frequency of unit root present in $\mathbf{y}_{t2}$. The trend parameter is also significant under 5% level. Nevertheless, It is clearer that $|\beta_1| = 1.9532 > 1$, $|\beta_2| = -1.2736 > 1$, $|\beta_3| = 0.3436 < 1$, $\frac{|\beta_1|}{|\beta_2|} = 1.5336 > 1$ and $\frac{|\beta_2|}{|\beta_3|} = 3.7066 > 1$. The values of $\beta_1$ and $\beta_2$ are significant under 1% and 5% level,
respectively. This result indicates that $y_2$ in the above (14) and (15) is I(2).

C. Unit Root Test Analysis

Here, few selected unit root tests results such as Augmented Dickey Fuller (ADF) test, Phillips-Perron (PP) Test; Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) Test; Elliott, Rothenberg, and Stock Point Optimal (ERS) Test and Ng and Perron (NP) Tests are being applied to compare the findings made using the new proposed AA(3) method of detecting integration order (frequency of unit root) in a data series.

The results of the ADF, PP, KPSS, ERS and NP unit root test in Table I – Table III indicate that the simulated series $y_1$ is integrated order one I(1) at the level series and I(0) or stationary at first difference and it is significant under 5% level. The empirical data on services sector output $y_2$ is integrated order two I(2) and stationary at the second difference. These results implied that simulated series $y_1$ has one unit root and $y_2$ has two unit roots.

### Table I: Analysis of Unit Root Test Using ADF and PP

| Test | Series | DT | Lags | Test Value | 1% level | 5% level | 10% level | Prob. | Rmk |
|------|--------|----|------|------------|----------|----------|----------|-------|-----|
| ADF  | $y_1$  | C, T | 22   | -3.1697    | -3.9627  | -3.1298  | -3.1279  | 0.0909 | I(1) |
|      |        |     |      |            | -3.1241 | -3.1280  | -3.1279  |        |     |
|      |        |     |      |            | -3.4268 | -3.4265  | -3.4250  |        |     |
|      | $\Delta y_1$ | C, T | 20   | -5.0720    | -3.5366  | -3.2003  | -3.2024  | 0.9946 | I(1) |
|      |        |     |      |            | -3.5403 | -3.5403  | -3.5403  |        |     |
|      |        |     |      |            | -2.9458 | -2.9458  | -2.9458  |        |     |
|      | $\Delta^2 y_1$ | C   | 4    | -5.1682    | -3.3627  | -3.3627  | -3.3627  | 0.0001 | I(0) |
|      |        |     |      |            | -3.0531 | -3.0531  | -3.0531  |        |     |
|      | $\nabla y_1$ | C, T | 22   | -3.1697    | -3.9627  | -3.1298  | -3.1279  | 0.0909 | I(1) |
|      |        |     |      |            | -3.1241 | -3.1280  | -3.1279  |        |     |
|      |        |     |      |            | -3.4268 | -3.4265  | -3.4250  |        |     |
|      | $\nabla^2 y_1$ | C   | 4    | -5.1682    | -3.3627  | -3.3627  | -3.3627  | 0.0001 | I(0) |
|      |        |     |      |            | -3.0531 | -3.0531  | -3.0531  |        |     |
|      | $\nabla^3 y_1$ | C   | 15   | -5.2650    | -3.2650  | -3.2650  | -3.2650  | 0.0001 | I(0) |
|      |        |     |      |            | -3.1983 | -3.1983  | -3.1983  |        |     |

### Table II: Analysis of Unit Root Test Using KPSS and ERS

| Test | Series | DT | Lags | Test Value | 1% level | 5% level | 10% level | Prob. | Rmk |
|------|--------|----|------|------------|----------|----------|----------|-------|-----|
| ADF  | $y_1$  | C, T | 34   | 0.6199     | 0.2160   | 0.1460   | 0.1190   | 0.2160 | I(1) |
|      |        |     |      |            | 0.2160   | 0.1460   | 0.1190   |        |     |
|      |        |     |      |            | 0.2160   | 0.1460   | 0.1190   |        |     |
|      | $\Delta^2 y_1$ | C   | 0    | 0.2033    | 0.4630   | 0.3470   | 0.3790   | 0.7390 | I(0) |
|      |        |     |      |            | 0.3470   | 0.3790   | 0.7390   |        |     |
|      | $\nabla^3 y_1$ | C   | 1    | 1.8700   | 6.1000   | 17.3000  | 14.2000  | 4.6959 | I(1) |
|      |        |     |      |            | 4.6959   | 14.2000  | 238.06  | 4.6959 | I(1) |

The $y_i$ represents the simulated data and $\gamma_i$ represents the empirical data (service sector output). The symbol $\nabla$ represents first difference and $\nabla^2$ indicates second difference.

D. Behaviour of AAR(3) Model in Log Transformed Series

In this section, pictorial representation of logarithm of and the simulated data $y_1$ empirical data (service sector output) $y_2$ are presented in Fig. 1 below.

The time series plots in Fig. 1 and 2 exhibit presences of drift in the log transformation of $y_1$ and $y_2$ and as such both series are non-stationary. The plots in Figure 3 and 4 exhibit a stationary outlook as the first log differencing transformation has eliminated the presence of drift in both variable $y_1$ and $y_2$. 
log(y_{i,t}) = 0.0197 + 1.87E-06 + 1.0886log(y_{i,t-1})
prob. (0.0000) (0.0000) (0.0000)
+ 0.0815log(y_{i,t-2}) - 0.1732log(y_{i,t-3}) + \epsilon_i
(0.0131) (0.0000)

(16)

The result in (12) indicates that the trend is significant and \( y_{i,t} \) is I(1) since \( |\beta_1| = 1.0886 > 1, |\beta_2| = 0.0815 < 1, |\beta_3| = -0.1732 < 1 \) and \( \frac{|\beta_1|}{|\beta_2|} = 13.3571 > 1 \) and the p-value (0.0000) of \( \beta_1 \) is significant under 5% level. This result reveals that there is only one unit root present in the simulated series \( \log(y_{i,t}) \). Therefore, first differencing can be used to achieve stationarity of \( \log(y_{i,t}) \). It was stated earlier that \( \beta_0 \) (the constant term) is optional. Not including the constant term does not affect the integration order of the variable under study. When there is no constant term, the result becomes:

\[
\log(y_{i,t}) = -2.58E-06 + 1.1988\log(y_{i,t-1})
\]
prob. (0.0000) (0.0000)
+ 0.0828\log(y_{i,t-2}) - 0.2810\log(y_{i,t-3}) + \epsilon_i
(0.0159) (0.0000)

(17)

This result also confirms the integration order of \( \log(y_{i,t}) \).

Giving that \( |\beta_1| = 1.1988 > 1, |\beta_2| = 0.0828 < 1, |\beta_3| = -0.2810 < 1 \) and \( \frac{|\beta_1|}{|\beta_2|} = 14.4783 > 1 \) the series \( \log(y_{i,t}) \) is I(1). And since the p-value (0.0000) is significant under 5% level, the hypothesis that \( H_0 : \beta_1 < 1 \) the \( \log(y_{i,t}) \) is I(0) is rejected in favour of the alternative. Hence, \( \log(y_{i,t}) \) is I(1).

The result of order of integration test for \( \log(y_2) \) is presented below:

\[
\log(y_{2,t}) = 0.3096 + 0.0126 + 1.5158\log(y_{2,t-1})
\]
prob. (0.0985) (0.3136) (0.0000)
- 0.4340\log(y_{2,t-2}) - 0.1481\log(y_{2,t-3}) + \epsilon_i
(0.1712) (0.4209)

(18)

The regression result in Equation (18) reveals that \( |\beta_1| = 1.5158 > 1, |\beta_2| = -0.4340 < 1, \) and \( |\beta_3| = 0.1481 < 1 \) \( \frac{|\beta_1|}{|\beta_2|} = 3.5138 > 1 \) \( \frac{|\beta_2|}{|\beta_3|} = 10.2350 > 1 \). The values of \( \beta_1 \) is significant under 5% level. This result entails that \( \log(y_{2,t}) \) is I(1). Conventionally, \( \log(y_{2,t}) \) can be stationary should the first logarithm differencing be applied. So, there is one unit root present in service sector output data. Omitting the constant term in (18), the result becomes as giving in (19) below:

E. Order of Integration Test Using AAR(3) Model in the Log Transformed Series

The results of the order of integration test for \( y_1 \) and \( y_2 \) via AAR(3) model are in Eq.(16) - (19) below:
The unit root results of the logarithm transformed variables $y_1$ and $y_2$ as presented in Table II. The ADF and PP unit root tests indicate that $\log(y_{1})$ is I(0) and $\log(y_{2})$ is I(2) and both are significant under 5% level. KPSS and NP report that $\log(y_{1})$ is I(2) significant at 5% and 10% respectively. For KPSS, $\log(y_{2})$ is I(0) significant under 5% level while NP reports that $\log(y_{2})$ is I(2) significant under 5% level. For ERS, $\log(y_{1})$ is I(1) significant under 5% level and $\log(y_{2})$ is I(2) significant under 5% level.

F. Unit Root Tests for the Log Transformed Variables

The section presents several unit root tests for the log transformed variables and the results are shown in Table IV and Table V below.

| Test    | Series | DT | Lags /BW | Test Value | 1% level | 5% level | 10% level | Prob | Rmk |
|---------|--------|----|----------|------------|----------|----------|-----------|------|-----|
| ADF     | $\log(y_{1})$ | C, T | 22 | -5.8803 | -3.9627 | 0.000 | I(0) |
|         | $\log(y_{1})$ | C, T | 23 | -3.1260 | -4.2191 | 0.000 | I(1) |
|         | $\log(y_{1})$ | C, T | 0  | -0.0867 | -3.1280 | -4.2191 | 0.000 | I(2) |
|         | $\log(y_{1})$ | C, T | 0  | -2.7117 | -3.5331 | -3.1280 | -4.2191 | 0.000 | I(3) |
|         | $\log(y_{1})$ | C, T | 0  | -7.5555 | -3.5366 | -3.1280 | -4.2191 | 0.000 | I(4) |
|         | $\log(y_{1})$ | C, T | 1  | 2.5756  | -3.1280 | -4.2191 | 0.000 | I(5) |
|         | $\log(y_{1})$ | C, T | 3  | -7.5555 | -3.1280 | -4.2191 | 0.000 | I(6) |
|         | $\log(y_{1})$ | C, T | 34| 0.9989  | -3.1280 | -4.2191 | 0.000 | I(7) |
|         | $\log(y_{1})$ | C, T | 26| 0.9714  | -3.1280 | -4.2191 | 0.000 | I(8) |
| PP      | $\log(y_{1})$ | C, T | 2  | -1.2576 | -3.1280 | -4.2191 | 0.000 | I(9) |
|         | $\log(y_{1})$ | C, T | 0  | -2.5756 | -3.1280 | -4.2191 | 0.000 | I(10) |
|         | $\log(y_{1})$ | C, T | 0  | -7.5555 | -3.1280 | -4.2191 | 0.000 | I(11) |
|         | $\log(y_{1})$ | C, T | 0  | -3.1280 | -4.2191 | 0.000 | I(12) |
|         | $\log(y_{1})$ | C, T | 0  | -3.1280 | -4.2191 | 0.000 | I(13) |
|         | $\log(y_{1})$ | C, T | 0  | -3.1280 | -4.2191 | 0.000 | I(14) |
|         | $\log(y_{1})$ | C, T | 0  | -3.1280 | -4.2191 | 0.000 | I(15) |

Note that ‘DT’ represents deterministic term and ‘Rmk’ represents remarks.
G. Discussion of Results

The results of the introduced order of integration test obtained in (12) and (13) revealed that the simulated data $y_1$ variable is I(1) or has one unit root. And the test results in (14) and (15) indicated that $y_2$ variable is I(2) or has two unit roots. These findings agree with the unit root test results of [3], [11], [6], [4], [1] and [10] as shown in Table I.

In the case of the logarithm transformed variables log($y_1$) and log($y_2$) as presented in Table IV and Table V. The introduced order of integration test using AAR(3) model in (16) and (17) revealed that Log($y_1$) is I(1). This finding agrees with that of ERS but disagrees with that of ADF and PP unit root tests that reported that log($y_1$) is I(0) and that of KPSS and NP whose results indicated that Log($y_1$) is I(2). For log($y_2$), AAR(3) results in (18) and (19) revealed that Log($y_2$) is I(1). This finding is contrary to that of ADF, PP, NP, ERS that reported I(2) and KPSS whose result indicated that Log($y_2$) is I(0).

The pictorial representation of log($y_1$) reveals more or less like S-shaped curve trend pattern and unlikely to be I(0) or stationary as claimed by [3], [11] and neither can it be I(2) as claimed by [6]. And the plot of first difference of natural logarithm of $y_2$ as shown in Fig. 1 reveals that V log($y_1$) is stationary or I(0), this confirms the results as presented in (16) and (17) that logarithm transformation of $y_1$ is I(1). This conspicuous reason implies that the introduced AAR(3) method of order of integration test performed better than other unit roots as exemplified in this paper.

IV. Conclusion

The paper presents a simple procedure for order of integration test that can be used as an alternative to unit root test. This procedure uses auxiliary autoregressive AAR(3) model framework to detect the order of integration for any given non-stationary variable. The generalized least square (GLS) was used to estimate the model parameters and hypotheses testing were based on t-statistic.

The order of integration test was applied to a simulated non-stationary series \{y_{it}\} with a sample size of 2000 and the result indicates that $y_{it}$ is integrated order one I(1). The test procedure was also applied to a known non-stationary variable $y_{2t}$ with two unit roots and the test detected the presence of two unit roots, in other words, $y_{2t}$ is integrated order two I(2).

The uniqueness of the test procedure lies in its simplicity within the univariate regression framework. The test procedure is apt for testing unit root and can compete favourably with any conventional unit root test.

REFERENCES

[1] Bhargava, A. On the theory of testing for unit roots in observed time series. Review of Economic Studies. 1986; 53: 369–384. Available: https://doi.org/10.2307/2297634.
[2] Central Bank of Nigeria. Statistical bulletin. 2020. Available: Central Bank of Nigeria | Annual Statistical Bulletin (cbn.gov.ng).
[3] Dickey, D.A. and W.A. Fuller. “Distribution of the Estimators for Autoregressive Time Series with a Unit Root,” Journal of the American Statistical Association, 1979; 74, 427–431. Available: https://doi.org/10.2307/2286348.
[4] Elliott, G., Rothenberg, T.J. and Stock, J.H. Efficient test for an autoregressive unit root. Econometrica, 1996; 64(4): 813-836. Available: http://dx.doi.org/10.2307/2171846.
[5] Harris, R and Sollis, R., Applied Time Series Modelling and Forecasting. Wiley and Sons Ltd, 2003.
[6] Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, Y. Shin. Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root, Journal of Econometrics, 1992; 54, pp. 159-178. Available: http: DOI: 10.12691/ijsmn-4-2-1.
[7] MacKinnon, James G. “Critical Values for Cointegration Tests,” Chapter 13 in R. F. Engle and C. W. J. Granger (eds.), Long-run Economic Relationships: Readings in Cointegration, Oxford: Oxford University Press, 1991.
[8] MacKinnon, James G. “Numerical Distribution Functions for Unit Root and Cointegration Tests,” Journal of Applied Econometrics, 1996; 11, 601-618. Available: DOI: 10.12691/jaams-4-5-1.
[9] Nelson, C.R., Plosser, C.I. Trends and random walks in macroeconomic time series. Journal of Monetary Economics, 1982; 10, 139–162. Available: http://dx.doi.org/10.1016/0304-3932(82)90012-5.
[10] Ng, S. and Perron, P. Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power. Econometrica, 2001; 69, 1519-1554. http://dx.doi.org/10.1111/1468-0262.00256.
[11] Phillips, P.C.B., P. Perron. Testing for a Unit Root in Time Series Regression. Biometrika, 1988; 75, pp. 335- 346. Available: DOI: 10.12691/jfe-4-2-3.
[12] Said, S.E. and Dickey, D.A. (1984) Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. Biometrika, 71, 599-607. Available: http://dx.doi.org/10.1093/biomet/71.3.599.
[13] Schwert, G.W. Tests for unit roots: A Monte Carlo investigation. Journal of Business and Economic Statistics,1989; 7, 147–159. Available: https://doi.org/10.2307/2291432.

Amaefula Chibuzo Gabriel is from Umuodo- Umukpanne in Ohaji-Egbema LGA of Imo state, Nigeria. He had B. Sc degree in statistics (1998) from Usmanu Danfodiyo University Sokoto, M. Sc degree in statistics(2002) from University of Benad and Ph. D in statistics(2014) from Usmanu Danfodiyo University Sokoto, all in Nigeria. He was a lecturer at the Federal Polytechnic Kaura-Namoda, Zamfara state, Nigeria (2008-2015), the head of Department of Mathematics and Statistics (2013-2015) and the Director of School of Information and Communication Technology (SICT) from 2013-2015. Presently, he is a senior lecturer in the Department of Mathematics and Statistics, Federal University Otuoke, Bayelsa state, Nigeria. He has published many articles in the field of time series and econometrics both in national and international Journals, Google Amaefula C.G for some information concerning previous publications.

Dr Amaefula C.G is a member of Nigerian Statistical Association (NSA) and International Research and Development Institute (IRDI).