International Journal of Statistics and Data Science, 2025, 1(1), 38-48
https://www.hillpublisher.com/journals/ijsds/
ISSN Online:

Solution to Double Differencing in Nonstationary Data Using a Difference Transformation Mechanism

Amaefula Chibuzo Gabriel[1,*], Eleje Chinwendu Best[2]

[1]Department of Mathematics and Statistics, Federal University Otuoke, Bayelsa State 562101, Nigeria.
[2]Department: Computer Science and Informatics, Federal University Otuoke, Bayelsa State 562101, Nigeria.

How to cite this paper: Amaefula Chibuzo Gabriel, Eleje Chinwendu Best. (2025) Solution to Double Differencing in Nonstationary Data Using a Difference Transformation Mechanism, *International Journal of Statistics and Data Science*, 1(1), 38-48.
DOI: 10.26855/ijsds.2025.12.004

Received: July 13, 2025
Accepted: August 1, 2025
Published: August 26, 2025

***Corresponding author:** Amaefula Chibuzo Gabriel, Department of Mathematics and Statistics, Federal University Otuoke, Bayelsa State 562101, Nigeria.

Abstract

This study introduces a novel difference transformation mechanism (DTM) designed to address the problem of double differencing in persistent unit root problem common in nonstationary time series data. The DTM eliminates double differencing when the underlying data series is integrated of order two (I(2)). This study utilizes three empirical non-stationary datasets: monthly observations of the All Share Index (ASI) spanning January 1985 to December 2023; annual data on total exports and re-exports of oil and non-oil products (TEXP) and services sector output (SSO) spanned from 1981 to 2020 and from 1981 to 2023 respectively. The auxiliary AR(3) order of integration test, alongside the ADF and PP unit root tests, were conducted on the three datasets. The analysis indicated that ASI is integrated of order one (I(1)), whereas TEXP and SSO are integrated order two (I(2)) in their original forms. Subsequently, the DTM was implemented on these variables, followed by a repetition of the integration and unit root tests on their transformed series. The results showed that the transformed variables ASI*, TEXP*, and SSO* attained stationarity significant at 5% level regardless of their initial orders of integration. These outcomes confirm the efficacy of the proposed DTM in achieving stationarity in non-stationary time series, irrespective of the number of unit roots present in the data variable. Consequently, DTM is recommended in analytical contexts where data stationarity is a critical prerequisite for robust statistical inference.

Keywords

Difference transformation mechanism; Non-stationary data; Persistent unit root

1. Introduction

In several fields of knowledge, data transformation techniques are required not only for improvement of data quality but also in making the insight extraction process more lucid. Many financial and time series data are non-stationary and do not always come in a structure that is directly appropriate for analysis. And many of such data are characterized with rising trend and regularly need transformation to achieve a stationary form before carrying out any analysis.

Many statistical test and models necessitate that data follow a normal distribution and the different groups of observations must come from populations which have a uniform variance or standard deviation. However, data transformations are applied to achieve specific objectives, such as ensuring linearity, attaining normality, or stabilizing variance. These transformations often necessitate fitting a linear regression model to the transformed variables rather than the original ones. This approach helps meet the assumptions underlying linear regression, leading to more reliable and interpretable results.

When data exhibits non-normality or heteroscedasticity, standard statistical tests may yield unreliable results. While ordinary least squares (OLS) estimators remain unbiased under such conditions, they are no longer the best linear unbiased estimators (BLUE), as their variance is no longer minimal. To address these issues, data transformations such as the Box-Cox or Yeo-Johnson methods are often employed to achieve normality and stabilize variance, thereby restoring the assumptions necessary for valid statistical inference. However, it's important to note that transformations can alter the relationship between variables, so their use should be carefully considered. Spurious regression results arise as a result of regressing a non-stationary explained variable on stationary or non-stationary predictors.

Many empirical researches in literature have produced misleading and unreliable results by undermining the essentials of data transformation or appropriate data transformation before statistical analysis. In some cases, after the first difference, there is still persistent unit root, indicating the need for a second difference. This is a common practice in time series and econometric modelling. However, finding a runoff solution to double differencing is the thrust behind the development of difference transformation mechanism for solving persistent unit root problem in non-stationary data.

One of the earliest transformation techniques is that of Box and Cox [1]. When variance instability is detected in a data series, transformations may be considered to achieve variance stability. However, some of these transformations such as cube, square, square root, cube root, or fourth root of the original series are commonly applied to stabilize variance and achieve normality. Among these, the natural logarithm is a widely used transformation that can stabilize variance, especially when data exhibits exponential growth or skewness. The Box–Cox transformation is a more general method that encompasses these and other power transformations, providing a systematic approach to identify the most appropriate transformation for a given dataset. This family of transformations, introduced by [1]., includes various power functions that can be applied to data to stabilize variance and make the data more closely approximate a normal distribution. Whether using the natural log or a specific power transformation, all these methods are considered part of the broader Box-Cox transformation family.

$$\left.\begin{aligned} x &= \frac{(x_t+\omega_0)^{\nu}-1}{\nu} && \text{if } 0<\nu<1 \\ x &= \operatorname{Ln} X_t + \omega_0 && \text{if } \nu = 0 \end{aligned}\right\} \tag{1}$$

Where, ω_0 represents a constant and ν is the transformation's shape parameter, with ν being any real number. When ν = 0.5, the Box–Cox transformation simplifies to a square root transformation. If the variance increases as the values of the series rise, choosing a ν value less than 1 is recommended to help stabilize variance. Conversely, if the variance decreases as the series values grow, then a ν value greater than 1 may be more appropriate. According to Pankratz [2], time series data must first be made stationary—through such transformations—before applying the Box-Jenkins modeling approach.

The study of Tukey [3] introduced a number of data transformation techniques, including log transformations and the use of boxplots in data exploration. Bickel and Doksum [4] observed that in structured models—such as transformed linear regression with small to moderate error variances—the asymptotic variances of parameter estimates are significantly larger when the transformation parameter ν is estimated from the data, compared to when it is known. Conversely, in unstructured models like transformed one-way analysis of variance with moderate to large error variances, the impact of estimating ν is less pronounced. Despite this, they cautioned that the performance of Box-Cox-type procedures is unstable and highly sensitive to model parameters in structured models with small to moderate error variances.

Generally, to achieve stationarity in time series data, differencing is typically applied to series exhibiting stochastic trends, while regression methods are used for those with deterministic trends. Nelson and Plosser [5] demonstrated that many macroeconomic time series are difference stationary, meaning their first differences are stationary. Enders [6] further elaborated on these concepts, emphasizing the importance of identifying the appropriate trend type to select the correct transformation method. Applying the wrong transformation can lead to inefficient estimations and misleading inferences. Renshaw and McCulloch [7] applied Box-Cox transformations to instrument calibration, highlighting data normalization in practical settings. Draper and Smith [8] in their book explore transformations in regression analysis and their impact on model fitting and interpretation.

In regression analysis, transformations are essential tools for addressing violations of standard model assumptions. Cook and Weisberg [9] emphasized that such transformations are pivotal when original variables exhibit non-constant mean and variance, which can undermine the validity of regression results. According to Samprit and Hadi [10], the assumptions most susceptible to violation include linearity and homoscedasticity, the latter referring to the constant variance of error terms.

Yeo and Johnson [11] introduced a flexible family of power transformations designed to improve normality or symmetry in data. Unlike the Box-Cox transformation, which is restricted to strictly positive values of x , their method accommodates both zero and negative values of x . The transformation is defined as follows:

$$\Phi(\nu,x)=\begin{cases}\dfrac{((x+1)^{\nu}-1)}{\nu} & \text{if } \nu\neq 0,\ x\geq 0\\ \mathrm{Ln}(x+1) & \text{if } \nu=0,\ x\geq 0\\ -[(-x+1)^{2-\nu}-1)] & \text{if } \nu\neq 2,\ x<0\\ -\mathrm{Ln}(-x+1) & \text{if } \nu=2,\ x<0\end{cases} \tag{2}$$

If x is strictly positive, the Yeo-Johnson transformation is equivalent to the Box-Cox transformation applied to (x+1). If x is strictly negative, the Yeo-Johnson transformation corresponds to the Box-Cox transformation applied to ($-x$+1), but with the transformation parameter ν adjusted to $2-\nu$. However, interpreting the Yeo-Johnson transformation parameter is challenging because it lacks a clear or intuitive connection to measurement units or magnitude. Its influence on the distribution varies based on the data's shape and skewness, making the transformation highly dependent on context and inherently nonlinear.

When a time series exhibits variance proportional to its mean or follows an exponential pattern, applying a logarithmic transformation can help stabilize the variance. This approach is particularly useful when the variability increases with the magnitude of the data, a common occurrence in various fields such as economics and biology [12].

Several study books explained the need of data transformation and some these are Jolliffe [13], who discussed the function of data transformations, such as centering and scaling, in Principal Component Analysis (PCA); Bishop [14] focused on data preprocessing, including transformation techniques like whitening and normalization, in pattern recognition and machine learning contexts; explored various data transformations—such as logarithmic, square root, and Box-Cox—to address issues like non-normality and heteroscedasticity in multivariate analysis. They emphasized the importance of selecting appropriate transformations to meet model assumptions and enhance the validity of statistical inferences. Their work provides guidance on applying these transformations effectively within multivariate statistical techniques. Wilcox [15] who explored robust statistical methods and data transformations to handle outliers and non-normality while Schumacker and Lomax [16] in their study provided insights on transformations in structural equation modeling (SEM), focusing on the scaling and normalization of data. Yin and Yi [17] in their work surveyed a variety of transformation methods such as scaling, encoding, and normalization in the context of machine learning and statistics.

The present paper offers a novel procedure to stabilizing variance for a known nonstationary series without lost of generality. Actually, it leaves many, if not all nonstationary data integrated order zero I(0) in a single transformation. The proposed transformation mechanism not only eliminates the problem of a persistent unit root in a data set but also eliminates double differencing when the underlying data variable has an integrated order two (I(2)) problem. Again, it removes trend and seasonality in time series data.

The modulus of the new method is an integrated differencing that is capable of rendering any nonstationary series stationary. Unlike conventional differencing where we have $n-1$ data points left after the first difference and $n-2$ data points after the second difference. The proposed transformation method reduces all non-stationary time series to $n-1$ observations, regardless of how many unit roots the data contains.

For example, in a linear regression setting where the independent variable is integrated of order two, I(2), it is standard practice to apply second-order differencing to achieve stationarity. If the dependent variable is integrated of order one (I(1)), a single differencing is typically sufficient to render it stationary. These will leave both the explained and the explanatory variables on different data points; $n-1$ and $n-2$ data points respectively. This scenario will course the regression estimates to produce a higher error variance, therefore, rendering the outcome of such analysis limited, if not spurious. The proposed transformation mechanism offers a significant advantage over traditional differencing methods by effectively rendering both integrated of order one (I(1)) and integrated of order two (I(2)) variables stationary. This capability diminishes the necessity for second-order differencing, thereby enhancing model efficiency and preserving the inherent characteristics of the original data.

2. Research Methods and Materials

This study employs an empirical research methodology, grounded in direct observation and experimentation to derive knowledge from actual experience rather than from theory or belief. The forthcoming section outlines the methodology of the proposed differencing mechanism, which is designed to transform a nonstationary variable into a stationary one, regardless of its order of integration.

2.1 Data Description, Sources and Methods of Collection

This study utilizes three distinct empirical datasets:

a) *All Share Index (ASI):* Monthly data spanning from January 1985 to December 2023.

b) *Services Sector Output (SSO):* Annual data covering the period from 1981 to 2023.

c) *Total Exports and Re-exports (TEXP)*: Annual data from 1981 to 2020.

Preliminary analysis indicates that the ASI data is integrated of order one (I(1)), while both the SSO and TEXP datasets exhibit integration of order two (I(2)) characteristics. The data utilized in this study were obtained from the National Bureau of Statistics (NBS) in Kaduna, in partnership with the Central Bank of Nigeria (CBN). Specifically, the ASI and SSO dataset are sourced from the CBN [18] Statistical Bulletin (2024 edition), while the TEXP datasets is obtained from the 2021 edition of the same publication. The ASI dataset comprises a total of 468 data points, whereas the SSO and TEXP datasets include 43 and 40 entries, respectively.

2.2 Integration Order Test: Test of Unit Root

Amaefula [19] highlights that the auxiliary autoregressive AAR(3) model provides an alternative approach for testing the presence of unit roots. In this study, both the Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) unit root tests are employed to reinforce the conclusion that the ASI, TEXP and SSO datasets are integrated of order one and two, respectively.

The Order of Integration Test (OIT) based on the AAR(3) model is shown in Equation (3) below.

$$\left.\begin{aligned} x_t - \sum_{j=1}^{3} \varphi_j x_{t-j} &= \tau_0 + \tau_1 trend + \eta_t \\ x_t - \sum_{j=1}^{3} \varphi_j x_{t-j} &= \tau_0 + \eta_t \\ x_t - \sum_{j=1}^{3} \varphi_j x_{t-j} &= \eta_t \end{aligned}\right\} \quad (3)$$

The initial model presented in Equation (3) is utilized when a trend is observed in the underlying data series. Conversely, the second model is appropriate when the data series shows no indication of a trend. The coefficient τ_1 represents the trend component, which may be statistically significant or not, depending on the data. The third model in (3) is applied with there is exclusion of the intercept τ_0 and the trend term. The stochastic error termand the φ_i's are the autoregression coefficients.

Test conditions are:

For I(1); $|\varphi_1| \geq 1, |\varphi_2| < 1, |\varphi_3| < 1$ and $\left|\frac{\varphi_1}{\varphi_2}\right| > 1$

For I(2); $|\varphi_1| > 1, |\varphi_2| \geq 1, \frac{|\varphi_1|}{|\varphi_2|} > 1$

In general, the null hypothesis is formulated as; H_0: $|\varphi_j|$is less than one against the alternative, H_1:that asserts at least one of the absolute value of φ_j's is greater than or equal to one.

Unit root test: ADF

The structure of the Augmented Dickey-Fuller (ADF) test can be expressed as follows:

$$\nabla y_t = \alpha + \alpha_1 t + \beta y_{t-1} + \sum_{j=1}^{k} \xi_j \nabla y_{t-j} + a_t \quad (4)$$

In this context, k denotes the number of lagged difference terms included in the model. Equation (4) incorporates an intercept, a drift component, and a deterministic trend. The ADF test evaluates the presence of a unit root, where the null hypothesis $H_0: \beta = 0$ and alternative $H_a: \beta$ is less than zero. As noted by Dickey and Fuller [20], if the computed ADF test statistic exceeds the critical values at the 1%, 5%, or 10% levels, the null hypothesis of a unit root cannot be rejected.

PP Test

The unit root test proposed by Phillips and Perron [21] is formulated as follows:

$$Z_\alpha = t_\alpha \left(\frac{v_0}{f_0}\right)^{\frac{1}{2}} - \frac{T(f_0 - v_0)(se(\hat{\alpha}))}{2f_0^{\frac{1}{2}}s} \tag{5}$$

Here, t_α represents the t-statistic for the estimate of α, while SE(α) denotes the standard error of the regression equation. Additionally, v_0 is a consistent estimator of the error variance, as outlined in Equation (6). It's important to note that $v_0 = (T-k)s^2/T$ and k is the number of explanatory variables, f_0 is an estimate of the spectral density of the residuals at frequency zero.

$$\nabla y_t = \alpha y_{t-1} + x_t{}'\delta + a_t \tag{6}$$

The optional exogenous variables x_t, can include a constant term alone or both a constant and a trend. The AR-based spectral estimator for the spectrum at frequency zero is then defined as

$$f_0 = \hat{\sigma}_a^2 / \left(1 - \hat{\beta}_1 - \hat{\beta}_2 - \cdots - \beta_p\right) \tag{7}$$

In (7), $\hat{a}_a^2 = \frac{\sum \hat{a}_t^2}{T}$ represents the variance of the residuals, and the β's are coefficients obtained from the auxiliary regression estimates.

2.3 Specification of the Differenced Transformation Mechanism (DTM)

An important benefit of the suggested transformation is its capacity to convert an integrated order one (I(1)) series into a stationary I(0) series, and additionally, to stabilize an integrated order two (I(2)) series without requiring a second differencing. And the specification is of the form;

Define the transformed variable as $\left[(x_t - x_{t-1})^2 / x^2{}_{t-1}\right] \times C$. Let, x_t^* denote the DTM variable so that,

$$\begin{aligned} x_t^* &= [(x_t - x_{t-1})^2 / x_{t-1}^2] \times C \\ &= [(x_t^2 + x_{t-1}^2 - 2x_t x_{t-1}) / x_{t-1}^2] \times C \\ &= \left[\frac{x_t^2 + x_{t-1}^2 - 2x_t x_{t-1}}{x_{t-1}^2}\right] \times C \\ &= \left[\frac{(x_t - x_{t-1})^2}{x_{t-1}^2}\right] \times C \end{aligned} \tag{8}$$

where C is a constant (e.g., 10, 10^2) chosen to scale-up the transformation. Without loss of generality, the algebraic expansion in (8) can be re-written as

$$x_t^* = \left[1 - \frac{x_t}{x_{t-1}}\right]^2 \times C \tag{9}$$

where the time series variable x_t is nonstationary, x_{t-1} denotes its first lag of x_t and C is the common constant (it can be 10, 10^2 etc) and its value does not change the stationary condition of the transformed variable. This method eliminates the need for differencing a series twice to attain stationary state. Note that (9) becomes our developed differenced transformation mechanism (DTM).

Let $\mathrm{x_t}$ be a non-stationary time series variable and let $\mathrm{x_t^*}$ be a transformed form of $\mathrm{x_t}$ by DTM. If $\mathrm{x_t}$ is integrated order one, then we have:

$$\mathrm{x_t = (x_1, x_2, \cdots, x_n) \sim I(1) \xrightarrow{DTM} x_t^* = (x_1, x_2, \cdots, x_{n-1}) \sim I(0)}$$

If x_t is integrated order two, we have

$$x_t = (x_1, x_2, \cdots, x_n) \sim I(2) \overset{DTM}{\rightarrow} x_t^* = (x_1, x_2, \cdots, x_{n-1}) \sim I(0)$$

If x_t is integrated order d (i.e, I(d)), we have

$$x_t = (x_1, x_2, \cdots, x_n) \sim I(d) \overset{DTM}{\rightarrow} x_t^* = (x_1, x_2, \cdots, x_{n-1}) \sim I(0)$$

Here d represents the order of integration or the number of times a non-stationary variable can be differenced to be stationary. Despite the order of integration of the transformed variable, we are left with $n-1$ observation points; this is the leverage of DTM over conventional differencing.

2.4 Estimation Method

Adopting the generalized linear method, Equation(1) may be expressed in matrix notation as;

$$x = X\beta + a \tag{10}$$

In (10), X denotes the vector containing lagged values of the dependent variable, $E(a) = 0$, var-cov$(a) = \sigma^2 V$and σ^2 is known. The matrix V, which is assumed to be known, characterizes the variance and covariance structure of the random error terms (a). Given this specification for the error variance-covariance, the β coefficients can be estimated using

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}x \tag{11}$$

The variance-covariance of $\hat{\beta}$ is equal to $\sigma^2(X'V^{-1}X)^{-1}$. When the error terms are assumed to have constant variance over time and are uncorrelated with one another, the V matrix simplifies to the identity matrix. However, in the presence of both heteroskedasticity and autocorrelation, V will contain non-zero elements along both its diagonal and off-diagonal entries.

3. Results and Discussion

This section presents the graphical data presentation of the variables (ASI, TEXP and SSO), order of integration test and the results from DTM application and their discussions.

3.1 Graphical Presentation of Data

The three empirical data; monthly all shares index (ASI), yearly total exports and re-export of oil and non oil products (TEXP) and yearly service sector output (SSO) are graphically presented in Figure 1 bellow.

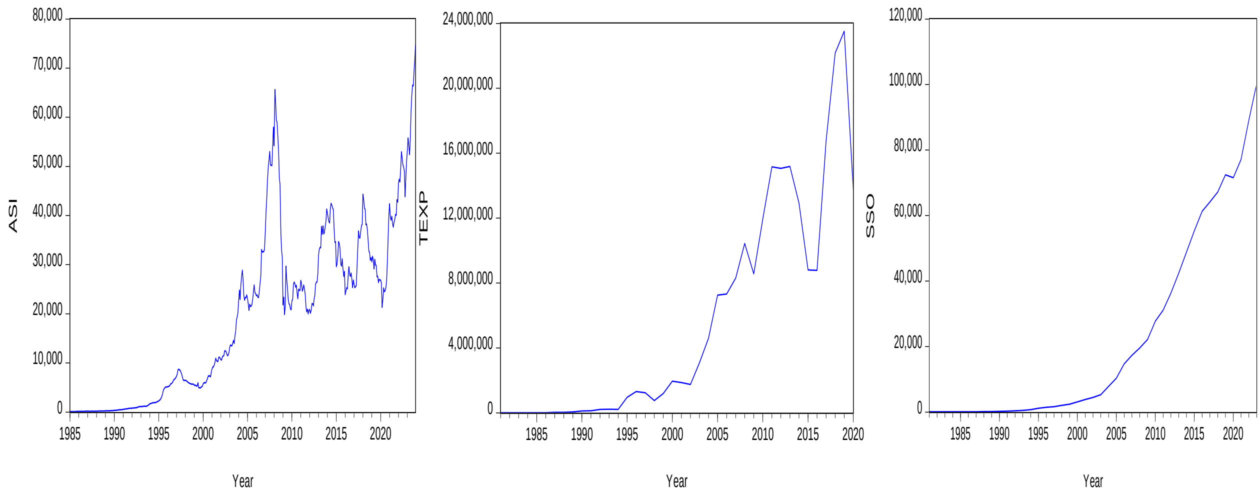

Figure 1. Time plots of ASI,(from 1985-2023) , TEXP(from 1981-2020) and SSO (from 1981-2023) in Nigeria.

The ASI shows greater fluctuations during the latter part of the series compared to the earlier part. ASI plot exhibits irregular drifts or fluctuations and positive trend. Nevertheless, the time plot shows a positive growth in ASI. The TEXP exhibits a steady rising pattern with notable drop in 2015 and 2020. Generally, there is evidence of a positive trend in TEXP variable. Variation in the ASI is more pronounced in the latter portion of the time series than in the initial segment. The Table1 below provides a more lucid description of the two variables under study.

Table1. Overview of Key Statistical Measures

Parameter	μ	σ	Sk	Kt
ASI	19449.89	17423.09	0.6508	2.6259
TEXP	5641582.0	6788913.0	1.0361	2.9977
SSO	22442.44	29668.99	1.1361	2.9388

Note: Sk =skewness and Kt = kurtosis

The descriptive overview in Table 1 indicates that the SSO distribution is more positively skewed and exhibits a heavier tail compared to TEXP. Additionally, the TEXP variable has higher values for both the mean and standard deviation relative to those of ASI and SSO The values of the kurtosis for the three variables are approximately 3 but their skewness show departure from normality and symmetry.

3.2 Estimates of AAR(3) Model Test

The analysis of the integration order, conducted using the auxiliary autoregressive AAR(3) model, is outlined below;

$$\mathrm{ASI_t} = -183.0122 + 2.7874\mathrm{TR} + 1.1373\mathrm{ASI_{t-1}} + 0.0018\mathrm{ASI_{t-2}} - 0.1577\mathrm{ASI_{t-3}} \quad (12)$$

In (12), ASI is classified as integrated of order one (I(1)) because the value of $|\varphi_1| > 1$ and the values of $|\varphi_2| < 1$ and $|\varphi_3| < 1$. This result indicates that ASI requires first difference to be stationary.

$$\nabla\mathrm{ASI_t} = -53.9781 + 0.6571\mathrm{TR} + 0.1228\nabla\mathrm{ASI_{t-1}} + 0.1216\nabla\mathrm{ASI_{t-2}} + 0.1576\nabla\mathrm{ASI_{t-3}} \quad (13)$$

Since all the absolute values of φ's are all less than one, the ∇ASI variable is integrated order zero (I(0)) or stationary.

$$TEXP_t = -2101481.0 + 201349.9TR + 1.3585TEXP_{t-1} - 1.1842TEXP_{t-2} + 0.4494TEXP_{t-3} \quad (14)$$

In (14), TEXP is classified as integrated of order two (I(2)) because the value of $|\varphi_1| > 1$, $|\varphi_2| > 1$ and $|\varphi_3| < 1$. This result indicates that TEXP variable need to be difference twice to be stationary.

$$\nabla^2 TEXP_t = 459641.1 - 23333.76TR - 0.1037\nabla^2 TEXP_{t-1} - 0.3750\nabla^2 TEXP_{t-2} - 0.6167\nabla^2 TEXP_{t-3} \quad (15)$$

Since in (15) $|\varphi_1| < 1$, $|\varphi_2| < 1$ and $|\varphi_3| < 1$ then $\nabla^2 TEXP_t$ is stationary.

$$\mathrm{SSO_t} = -1096.009 + 107.7691\mathrm{TR} + 1.4930\mathrm{SSO_{t-1}} - 0.7372\mathrm{SSO_{t-2}} + 0.2695\mathrm{SSO_{t-3}} \quad (16)$$

In (16), SSO variable data exhibits of a persistent unit root problem as $|\varphi_1| = 1.4930 > 1$, $|\varphi_2| \to 1$ and $|\varphi_3| < 1$. This result indicates that SSO variable data has a near second unit root problem as the absolute value of ϕ_2 tends to unity. and it's considered as integrated order two (I(2)), hence, requires second differencing to attain a stationary state.

$$\nabla^2\mathrm{SSO_t} = -531.0583 + 38.2717 - 0.2258\nabla^2\mathrm{SSO_{t-1}} - 0.26590\nabla^2\mathrm{SSO_{t-2}} + 0.1902\nabla^2\mathrm{SSO_{t-3}} \quad (17)$$

Since in (17), $|\varphi_1| < 1$, $|\varphi_2| < 1$ and $|\varphi_3| < 1$ then $\nabla^2 SSO_t$ is stationary.

3.3 Analysis of Unit Roots: ADF and PP Tests

Table 2 below provides a summary of the ADF and PP unit root test outcomes.

According to the ADF, the ASI variable exhibits a single unit root problem and its stationary after first difference, TEXP exhibits double unit root problem while SSO exhibits persistent unit root problem even after second difference. But based on PP unit root tests presented in Table 2 ASI is integrated order one, TEXP exhibits weak stationarity, significant at 10% level after the second difference and SSO is integrated order two (I(2)). This implies that ASI is integrated of order one whereas TEXP and SSO are integrated of order two (I(2)) according to PP unit root test and these results tally with that of AAR(3) OIT in sub-section 3.2.

Table 2. Overview of ADF and PP Unit Root Test Results

Test Type	Parameter	DT	Band-Width/Lag	Test value	Critical values 0.01 0.05 0.10	Prob*	Remark
ADF	ASI	Constant & Trend	3	-2.6029	-3.9780 -3.4195 -3.1324	0.2792	Not stationary at the level series, first difference indicating I(1)
	∇ASI	Constant & Trend	3	-9.1735	-3.9780 -3.4195 -3.1324	0.0000	Stationary under 5% after first difference
	TEXP	Constant & Trend	7	-3.2097	-4.2733 -3.5577 -3.2124	0.1005	Not stationary at the level series, first difference indicating I(1)
	$\nabla TEXP$	Constant & Trend	9	-1.9842	-4.3098 -3.5742 -3.2217	0.5852	Not stationary after first difference indicating I(2)
	$\nabla^2 TEXP$	Constant & Trend	5	-6.5908	-4.2733 -3.5577 -3.2124	0.0000	stationary under 5% after second difference
	SSO	Constant & Trend	9	-1.2056	-4.2627 -3.5530 -3.2096	0.8929	Not stationary at the level series, first difference indicating I(1)
	∇SSO	Constant & Trend	9	-1.9316	-3.9778 -3.4195 -3.1323	0.6149	Not stationary after first difference, indicating I(2)
	$\nabla^2 SSO$	Constant & Trend	9	-2.9461	-4.2846 -3.5629 -3.2153	0.1628	Not stationary after second difference, indicating persistent unit root
PP	ASI	Constant & Trend	10	-2.2413	-3.9778 -3.4195 -3.1323	0.4649	Not stationary at the level series, first difference indicating I(1)
	∇ASI	Constant & Trend	9	-18.9997	-3.9778 -3.4195 -3.1323	0.0000	Stationary under 5% after first difference
	TEXP	Constant & Trend	23	-2.2100	-4.2119 -3.3597 -3.1964	0.4710	Not stationary at the level series, first difference indicating I(1)
	$\nabla TEXP$	Constant & Trend	37	2.0576	-4.2191 -3.5331 -3.1983	1.0000	Not stationary after first difference, first difference indicating I(2)
	$\nabla^2 TEXP$	Constant & Trend	5	-3.4654	-4.2268 -3.5366 -3.2003	0.0582	Stationary at 10% after second difference first difference, indicating weak stationarity
	SSO	Constant & Trend	0	1.9505	-4.1923 -3.5208 -3.1913	1.0000	Not stationary at the level series, first difference indicating I(1)
	∇SSO	Constant & Trend	1	-3.4359	-4.1985 -3.5236 -3.1929	0.0605	Not stationary under 5% level after first difference, first difference indicating I(2)
	$\nabla^2 SSO$	Constant & Trend	0	-7.5180	-4.2050 -3.5266 -3.1946	0.0000	stationary under 5% after second difference

DT denotes determinist term; T-value represents test value

3.4 Application of the Developed DTM

Having identified the order of integration in ASI, TEXP and SSO variables under sections 3.2-3.3, we will apply the developed transformation mechanism (DTM) to the ASI, TEXP and SSO variables and repeat the order of integration test.

Time plot of the Transformed variables

In this section, we utilized the established DTM framework on the ASI, TEXP, and SSO datasets, with the resulting time series of the transformed variables illustrated in Figure 2. It is important to mention that a single application of the DTM was performed on the ASI, TEXP, and SSO data.

The time plots in Figure 2 above exhibit the application of the DTM to the original ASI, TEXP and SSO data. The concentration of the transformed ASI*, TEXP* and SSO* around zero with no trend component signifies that the means are approximately zero and the stability of the variances. It also implies that the three variables are stationary irrespective of the nature of integration of their original data.

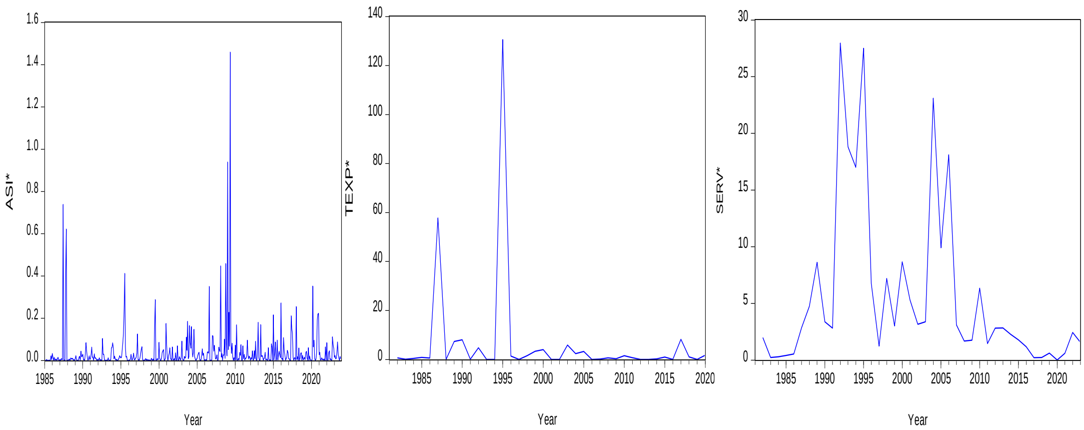

Figure 2. Time plot of ASI, TEXP and SSO transformed using the DTM.

3.5 Testing the Integration Order of the Transformed Variables

The integration order analysis in Section 4.2, along with the unit root assessments in Section 4.3, indicated that ASI is integrated of order one (I(1)), while TEXP and SSO exhibit integration of order two (I(2)). It is anticipated, however, that applying these tests to the transformed variables ASI*, TEXP*, and SSO* will reveal they are stationary, or integrated of order zero (I(0)). The corresponding outcomes are presented in Equations (18)-(20).

$$ASI_t^* = 0.3170 + 0.0959ASI_{t-1}^* + 0.0780ASI_{t-2}^* + 0.0526ASI_{t-3}^* \quad (18)$$

In (18), the values of all ϕ coefficients are below one, that is, hence, ASI* is determined to be integrated of order zero (I(0)). This implies that ASI* variable is stationary. Therefore, we can conveniently assert that the application of the DTM has rendered an integrated order one ASI variable stationary.

$$TEXP_t^* = 80.5054 - 0.0779TEXP_{t-1}^* - 0.0671TEXP_{t-2}^* - 0.0246TEXP_{t-3}^* \quad (19)$$

Since $|\varphi_1| < 1$, $|\varphi_2| < 1$and $|\varphi_3| < 1$ in (19) then, TEXP* is stationary. Therefore, we can conveniently assert that a single application of the developed DTM has rendered an integrated order two *TEXP* variable stationary. This is the advantage of DTM over ordinary differencing stationary.

$$SSO_t^* = 8.6041\text{-}6.09\text{E-}07SSO_{t-1}^* - 0.0006SSO_{t-2}^* + 0.0005SSO_{t-3}^* \quad (20)$$

Therefore, Since $|\varphi_1| < 1$, $|\varphi_2| < 1$and $|\varphi_3| < 1$ in (20) then, SSO* is integrated order zero (I(0)) or we can conveniently say that a single application of the developed DTM has rendered an integrated order two SSO variable stationary.

3.6 Validation Test Analysis

To confirm the findings presented in sub-section 3.5, the transformed variables ASI*, TEXP*, and SSO* are also subjected to the ADF and PP unit root tests. A summary of the test outcomes is provided in Table 3 below.

Table 3. Overview of Unit Root Test Results for the Transformed Variables ASI*, TEXP*, and SSO*

Test Type	Parameter	DT	Band-width /Lag	Test value	Critical values 0.01 0.05 0.10	Prob*	Remark
ADF	ASI*	Constant	3	-7.4689	-34442 -2.8676 -2.5700	0.0000	Stationary under 5%
	TEXP*	Constant	0	-6.4190	-3.6156 -2.9411 -2.6091	0.0000	Stationary under 5%
	SSO*	Constant	0	-3.7324	-3.6010 -2.9350 -2.6058	0.0071	Stationary under 5%
PP	ASI*	Constant	12	-20.4335	-34442 -2.8676 -2.5700	0.0000	Stationary under 5%
	TEXP*	Constant	2	-6.4285	-3.6156 -2.9411 -2.6091	0.0000	Stationary under 5%
	SSO*	Constant	3	-3.7929	-3.6010 -2.9350 -2.6058	0.0060	Stationary under 5%

DT denotes determinist term; T-value represents test value

In Table 3, the outcomes of the unit root tests confirm that the transformed variables ASI*, TEXP*, and SSO* are stationary, indicating they are integrated of order zero (I(0)). This result supports the earlier integration test findings discussed in subsection 3.4.2. Consequently, it can be conclusively stated that applying the DTM to the ASI, TEXP, and SSO series effectively renders them stationary, regardless of the initial number of unit roots present.

3.7 Discussion of Results

The three empirical variables (ASI, TEXP, and SSO) were analyzed using the AAR(3) integration test. The results indicated that ASI is integrated of order one (I(1)), while TEXP and SSO are integrated of order two (I(2)). To confirm these integration orders, the ADF and PP unit root tests were also conducted, and their outcomes were consistent with those obtained from the AAR(3) test.

The differenced transformation mechanism (DTM) was applied to the ASI, TEXP and SSO variables and we subjected the transformed data variables ASI*, TEXP* and SSO* to the same order of integration test in subsection 3.4.2, the result showed that the variables ASI*, TEXP* and SSO* are stationary or I(0). The validation test using ADF and PP unit root test in Table3 confirm same.

4. Conclusion

The empirical application of the developed DTM revealed that the mechanism is effective in making an integrated order two data variable stationary or integrated order zero (I(0)). Thus, it can be conclusively asserted that the DTM approach possesses the capability to eliminate the need for second differencing in instances where nonstationary data series exhibit integrated order two problem or persistent unit root behavior. However, the DTM approach is recommended in scenarios where achieving stationarity is a critical prerequisite for subsequent data analysis.

Acknowledgement

We acknowledge the support of the University administration in securing TETFUND approval for the sponsorship of this institutional-based research (Ref: FUO/AP/TETF.24IBR/001). We also extend our appreciation to the National Bureau of Statistics for providing access to the datasets utilized in this study.

References

[1] Box GEP, Cox DR. An analysis of transformations. J R Stat Soc Ser B. 1964;26:211-52.

[2] Pankratz A. Forecasting with dynamic regression models. New York: Wiley; 1991. ISBN: 0471615285.

[3] Tukey JW. Exploratory data analysis. Reading (MA): Addison-Wesley; 1977.

[4] Bickel PJ, Doksum KA. An analysis of transformations revisited. J Am Stat Assoc. 1981;76(374):296-311. doi:10.1080/01621459.1981.10477649.

[5] Nelson CR, Plosser CI. Trends and random walks in macroeconomic time series. J Monet Econ. 1982;10:139-62.

[6] Enders W. Applied econometric time series. New York: Wiley; 1995.

[7] Renshaw AE, McCulloch RE. Application of the Box-Cox transformation to the calibration of analytical instruments. Technometrics. 1996;38(1):69-74.

[8] Draper NR, Smith H. Applied regression analysis. 3rd ed. New York: Wiley; 1998.

[9] Cook RD, Weisberg S. Applied regression including computing and graphics. New York: Wiley; 1999.

[10] Chatterjee S, Hadi AS. Regression analysis by example. 4th ed. Hoboken (NJ): Wiley; 2006.

[11] Yeo IK, Johnson RA. A new family of power transformations to improve normality or symmetry. Biometrika. 2000;87(4):954-9. doi:10.1093/biomet/87.4.954.

[12] Yaffee RA, McGee M. Introduction to time series analysis and forecasting with applications of SAS and SPSS. San Diego: Academic Press; 2000.

[13] Jolliffe IT. Principal component analysis. 2nd ed. New York: Springer; 2002.

[14] Bishop CM. Pattern recognition and machine learning. New York: Springer; 2006.

[15] Wilcox RR. Introduction to robust estimation and hypothesis testing. 3rd ed. San Diego: Academic Press; 2012.

[16] Schumacker RE, Lomax RG. A beginner's guide to structural equation modeling. 4th ed. New York: Routledge; 2016.

[17] Yin X, Yi F. Data transformation methods for data preprocessing in machine learning: A survey. J Comput Sci Technol. 2018;33(1):89-102.

[18] Central Bank of Nigeria. Statistics bulletin [Internet]. 2024 [cited 2024 Jun 19]. Available from: https://www.cbn.gov.ng.

[19] Amaefula CG. A simple integration order test: An alternative to unit root testing. Eur J Math Stat. 2021;2(3):77-85. doi:10.24018/ejmath.2021.2.3.22.

[20] Dickey DA, Fuller WA. Distribution of the estimators for autoregressive time series with a unit root. J Am Stat Assoc. 1979;74:427-31.

[21] Phillips PCB, Perron P. Testing for a unit root in time series regression. Biometrika. 1988;75:335-46.