Empirical Similarity-Based Approach for Selection of Unit Root Test

Jude Kwesi Gyan Intsiful [1], Bismark Kwao Nkansah[1], David Kwamena Mensah[1*], Francis Eyiah-Bediako[1]

1. Department of Statistics, University of Cape Coast, Cape Coast, Ghana

* E-mail of the corresponding author: dmensah@ucc.edu.gh

Abstract

The existence of unit roots in time series processes can impair the choice of techniques for analysis and forecasting time series data. It is of much importance in econometric modelling to determine the integration number of analyzed time series based on unit root tests. Though statistical theory provides broad range of unit root tests in standard softwares, the choice of an appropriate test highly depends on subjective assessment of the analyst. This paper considers similarity-based scoring approach for selecting the most appropriate unit root test for specific type of time series observations based on Chi-square statistic and which is able to reduce subjectivity. Six unit root tests are studied. The utility of the proposed method is illustrated in simulation. The most reliable test, which is found is applied to a real time series of some selected macroeconomic variables.

Keywords: Time series, Stationarity, Unit root, Integration order, Chi square statistic

1. Introduction

The presence of unit roots in time series process can be informative on the choice of appropriate analysis techniques for the data as the construction of analytic models is based on the fundamental requirement of time series stationarity. As a result, the application of unit root tests for stationarity assessment of time series process has gained much recognition recently within the mainstream statistics. A practical application of unit root tests is seen in crop yields. Zapata et al. (2011) for example investigated non-stationarity in corn and soybean yields in the Delta State based on county-level data, using Dickey-Fuller test (Dickey and Fuller, 1979) and Phillip Perron test (Phillips and Perron, 1988) and their modification. Also, in macroeconomic modelling, unit root test has been the state-of-the-art approach for determining the most appropriate integration order of analyzed time series in order to identify some important features, e.g., stationarity.

There is a large literature on unit root tests for time series data that discuss the various unit root testing procedures that statistical theory provides. For example, Dickey and Fuller (1979), Phillips and Perron (1988), Schmidt and Phillips (1992), Elliott et al. (1996) and Kwiatkowski et al. (1992) handle the fundamental properties of the most commonly used tests, namely; Dickey-Fuller test, Augmented Dickey-Fuller test, Augmented Dickey-Fuller Generalized Least Square test, Phillip-Perron test, Schmidt and Perron and Kwiatkowski, Phillips, Schmidt and Shin test. Though these tests are formally integrated into standard statistical and econometric software's as well as other time series analytic packages, there hardly exist a software based approach for performing test selection in terms of selecting the most suitable test. Interestingly, analysts have been left to the mercy of subjective judgment in choosing an appropriate test. Although this subjective approach could be advantageous to experienced analysts, it could be expensive especially for large series, in that it will certainly require lots of efforts as well as much prior analysis and comparison work. This has created a great deal of interest in flexible methods for choosing the most reliable unit root test recently. Fedorova (2016) proposed an approach for performing unit root tests selection based on length of the series and value of the autoregressive parameter, using test power as the main performance measure.

Conventionally, assessment of statistical hypothesis has been by computation of the power which is the probability of rejecting a false null hypothesis (i.e., Type II error probability). Nevertheless, power calculations are based on knowledge of the alternative distribution of the test statistic under consideration. Due to this, in some situations the power cannot be computed. For hypothesis test where the alternative distribution of the test statistic is not available in closed form or challenging to derive, the power calculations comes with some computational challenge. This is the case of unit root testing. In particular, in unit root test hypothesis, the null distribution of the test statistic is not readily available in closed form. As a result, work in this field has been centered on simulation approach to power calculation (see, for example Fedorova, 2016). Though this approach is appealing since it addresses the use of subjectivity, it has some drawback in that the approach is dependent on knowledge of the autoregression (AR) parameter value which requires estimation via model fitting. This raises a simple question; should an analyst know the value of AR parameter before applying any test? To address this challenge, we exploit power calculations to develop an empirical similarity-based approach based on chi-square statistic leveraging on the values of the AR parameter and which addresses the strict dependence of existing method on the AR parameter value. This approach is simple to implement and less expensive when an appropriate model has not been agreed upon and one would like to explore several models.

The rest of the paper is structured as follows. The selected unit root tests are reviewed in brief in Section 2. In Section 3 we formulate and present our proposal. Application of the proposed method to simulated time series and real datasets are given in Section 4. Conclusions are drawn in Section 5.

2. Selected Unit Root tests

2.1 Dickey-Fuller Test

Consider the following data generating models satisfying first order autoregressive statistical models (Box and Jenkins, 1970)

$$x_t = \theta x_{t-1} + v_t, \quad t = 1,\ldots,T \tag{1}$$

$$x_t = a_0 + \theta x_{t-1} + v_t, \quad t = 1,\ldots,T \tag{2}$$

$$x_t = a_0 + a_1 t + \theta x_{t-1} + v_t, \quad t = 1,\ldots,T\ , \tag{3}$$

where θ denotes the autoregressive parameter, v_t represents the non-systematic component of the model that follows a white noise process.

Let $\hat{\theta}$ denote the least square estimator of θ and write $\sigma(\beta)$ for the standard error of the estimator β. Test of hypothesis of unit root based on models (1), (2) and (3) is termed the Dickey-Fuller test. The test employs hypothesis of the form

$$H_0 : \theta = 1, \text{ vrs } \ H_1 : |\theta| < 1$$

with standard test statistic

$$t_{DF} = \frac{D(\hat{\theta})}{\sigma(\hat{\theta})}, \tag{4}$$

where $D(\hat{\theta}) = \hat{\theta} - 1$. If the null hypothesis is true, the process contains a unit root and is characterized as a non-stationary process, written in standard notation as $I(0)$. The alternative hypothesis implies that there is no unit root present in the process and thus the process is characterised as stationary, written as $I(1)$. The representation $I(a)$ denotes an integrated process of order a. The null distribution of the test statistic (4) is the Dickey-Fuller distribution. Thus, hypothesis decisions are based on simulated critical values obtained by Dickey (1976) and Fuller (1976).

2.2 Augmented Dickey-Fuller Test

The possibility of autocorrelated error terms in the AR process cannot be ruled out in time aeries analysis. Suppose the error terms v_t are autocorrelated. Then model (1) can be expressed as

$$x_t = \theta x_{t-1} + \sum_{i=1}^{q-1} \varphi_i \tilde{x}_{t-i} + v_t, \quad \tilde{x}_{t-i} = \Delta x_{t-i} \tag{5}$$

where q is the lag length and ϕ_i are the AR coefficients corresponding to Δx_{t-i} for $i = 1,\ldots,q-1$. It is easy to see that Model (5) is a simple augmentation of the autoregressive unit root test model (1) and follows an ARMA structure. Test of hypothesis that a given time series is an $I(1)$ process against the alternative that it is an $I(0)$ process with the fundamental assumption that the data is characterised by an ARMA structure is termed

Augmented Dickey-Fuller (ADF) test (Dickey & Fuller, 1981). The computation of the test statistics employs the model

$$\Delta x_t = (\theta - 1)x_{t-1} + \sum_{i-1}^{q-1} \phi_i \tilde{x}_{t-i} + v_t \tag{6}$$

The corresponding constant or trend model based on (5) can be written as

$$x_t = \sum_{i=0}^{1} a_i t^i + \theta x_{t-1} + \sum_{i=1}^{q-1} \phi_i \tilde{x}_{t-i} + v_t, \quad \tilde{x}_{t-i} = \Delta x_{t-i} \tag{7}$$

The ADF test statistic has limiting distribution that is identical to the distribution of the DF test statistic. An imperative practical concern for ADF test is the specification of the lag length, q. It is well known that too small q leads to bias test due to the remaining serial correlation in the errors and too large q results in a test with low power. The choice of appropriate q has attracted attention in the literature. Ng and Perron (1995) proposed an empirical approach for choosing q that results in a stable test size, leading to minimal power loss. Their proposal involves first bounding q by an upper bound, say, $q_{\max}$ and then applying it in ADF test regression. If the absolute value of the *t*-statistic for assessing the significance of the last lagged difference is larger than 1.6, then q is set to $q_{\max}$ and unit root test is performed. Otherwise, q is set to $q_{\max} - 1$ and the process is repeated. Schwert (2002) proposed a useful rule of thumb for choosing $q_{\max}$ based on the length of the time series using

$$q_{\max} = \left[\left(\frac{T}{100} \right) \right]^{\frac{1}{4}},$$

where $[x]$ is the integer part of x. Assuming an ARMA process of unknown order for the errors, ensures the validity of the ADF test regression since the above choice for the lag length allows $q_{\max}$ to grow with the length of the time series under consideration.

2.3 Phillip-Perron Test

Phillip-Perron (PP) test is an attractive alternative to ADF test that addresses asymptotic bias appearing in the original ADF test when the residuals are characterised by serial correlation. Phillips and Perron (1988) considered nonparametric test statistic for testing the null hypothesis of unit root that permits weak dependence and heteroscedastic error process. The PP test can be viewed as a generalization of the ADF test based non-parametric correction to the Dickey-Fuller test statistic. Let V_T and $\hat{\varpi}_{i,T}$, denote, respectively, the ordinary least square (OLS) estimator of the non-systematic component and maximum likelihood estimator of the non-systematic covariance

$$V_T = \sum_{t=1}^{T} \frac{v_t^2}{T}, \quad \hat{\varpi}_{i,T} = T^{-1} \sum_{t=i+1}^{T} \hat{v}_t \hat{v}_{t-1}.$$

Define the statistic

$$V_{LT}^2 = V_T^2 + 2\sum_{i=1}^{T} \left(1 - \frac{i}{q+1} \right) \varpi_{i,T}.$$

For models involving constant terms, Pesaran (2015) define, test statistics for PP test of the form

$$Z_T = R(V_T, V_{LT})\, t_{DF} - \frac{1}{2} D^2(V_{LT}, V_T)\ R(T, V_{LT})\, R(V_{\hat{\theta}}, V_T) \tag{8}$$

and

$$Z_\theta = T(\hat{\theta}_T - 1) - \frac{1}{2}T^2 R^2(V_{\hat{\theta}}, V_T)\ D(V_{LT}, V_T), \tag{9}$$

where we have set $V_{\hat{\theta}} = \sigma(\hat{\theta})$ and defined $R(x_1, x_2) = \frac{x_1}{x_2}$, $D(x_1, x_2) = x_1 - x_2$ so that $R^2(x_1, x_2) = \frac{x_1^2}{x_2^2}$ and $D^2(x_1, x_2) = x_1^2 - x_2^2$, respectively. Given that the error terms are not autocorrelated, we have $V_{LT} = V_T$, $\hat{\varpi}_{i,T} = 0$, $i > 0$. Thus, the limiting distribution of the *t*- statistic is free of the autocorrelated parameters of the error process. In this case, the statistic (8) and (9) are exactly the Dickey-Fuller test statistic (t_{DF}). Though, PP and ADF tests are asymptotically equivalent, PP test exhibits better small sample properties than ADF.

2.4 Augmented Dickey-Fuller Generalized Least Square Test

Another modification of ADF test based on detrended transformation with generalized least square estimation method is termed Augmented Dickey-Fuller Generalized Least Square (ADF-GLS) test (Elliott et al., 1996). Consider the constant model obtained from Model (7) by setting $i = 0$ in the first term, which gives

$$x_t = a_0 + \theta x_{t-1} + v_t \tag{10}$$

Consider the transformation:

$$x_1' = x_1, \quad x_t' = x_t - \eta x_{t-1} \tag{11}$$

$$z_1 = 1, \quad z_t + \eta = 1, \quad T(1-\eta) = 7, \tag{12}$$

where $t = 1, \ldots, T$.

Based on the transformed model

$$x_t' = a_0 z_t + v_t, \tag{13}$$

a least square estimator of a_0, say $\hat{a}_0$, is obtained and applied to construct new time series of the form

$$x_t^* = x_t - \hat{a}_0 .$$

The final step involves the application of ADF test on the transformed time series based on

$$\Delta x_t^* = \theta x_{t-1} + \sum_{i=1}^{q-1} \phi_i \Delta x_{t-i}^* + v_t .$$

In the case of linear trend models, detrending process follows the steps outlined above with a modified transformation. In particular, transformations (11) and (12) are utilized together with

$$w_1 = 1, \quad w_t = t(1-\eta) + \eta, \ T(\eta - 1) = -13.5. \tag{14}$$

Hypothesis test decision for this test is based on simulated critical values tabulated by Elliott et al. (1996).

2.5 Schmidt and Phillips Test

Schmidt and Phillips (1992) developed likelihood maximization (LM) principle based approach to handle the drawback of DF-type tests and called their test Schmidt and Phillips (SP) test. Consider the statistical model

$$x_t = \beta_0 + z_t'\beta + y_t, \quad y_t = \mu y_{t-1} + v_t, \tag{15}$$

where $v_t \sim iid\ N(0, \sigma_0^2)$ and $z_t = (t, \ldots, t^q)$. Employing the idea of ordinary least squares, β_0 is estimated using say $\hat{\beta}_0$ via the model

$$\Delta x_t = \beta_0' \Delta z_t + \varepsilon_t \tag{16}$$

and applied to transform the underlying time series based on the statistic $\psi(x)$ defined as follows:

$$\psi(x) = x_1 - z_t'\hat{\beta}_0$$

The new series is constructed using the model

$$\tilde{x}_t = x_t - \psi(x) - z_t'\hat{\beta}_0$$

The test is then constructed based on the regression model

$$\Delta x_t = \lambda' \Delta z_t + \theta \tilde{x}_{t-1} + v_t \tag{17}$$

with the standard test statistic defined as

$$Z = \frac{T\hat{\theta}}{d(\hat{\theta})},$$

where $d(\hat{\theta}) = \left[1 + 2\frac{\sum_{k=1}^{l}\sum_{t=k+1}^{T} \hat{v}_t^2 v_{t-k}^2}{\sum_t \hat{v}_t^2} \right]^{-1}$. The SP test also tests the null hypothesis that a given series follows an integrated process of order one, i.e., $I(1)$ against the alternative that the series is $I(0)$.

2.6 Kwiatkowski, Phillips, Schmidt and Shin Test

Consider time series data that can be decomposed as

$$x_t = \sum_{i=0}^{q} a_i t^i + \mu_t + v_t,\ q = 0,\ 1,\ t = 1,\ldots,T \tag{18}$$

$$\mu_t = \mu_{t-1} + \varepsilon_t,$$

where μ_t follows a random walk and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. Consider testing a claim based on the variance of the random walk process so that the hypothesis of interest becomes

$$H_0 : \sigma_\varepsilon^2 = 0; \quad H_1 : \sigma_\varepsilon^2 > 0 \tag{19}$$

Unit root testing procedure based on (18) and (19) with the test statistic

$$LM = \frac{\sum_{t=1}^{T} S_t^2}{\hat{\sigma}_v^2},\quad S_t = \sum_{j=1}^{t} v_j,\quad t = 1, 2, \ldots, T$$

is known as Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test (Kwiatkowski et al., 1992). Here, S_t is the partial sum process of the residuals from a simple linear regression of x_t on an intercept and time, $\hat{\sigma}_v^2$, denotes the estimate of σ_ε^2 of the process v_t from (18). Test decisions are based on critical values tabulated by Kwiatkowski et al. (1992).

3 Similarity-based Scoring Scheme

We exploit hypothesis power computation based on simulations in this section to develop a new approach for addressing unit root test selection. Consider test of hypothesis of the form

$$H_0 : \theta = 1; \quad H_1 : |\theta| < 1.$$

Based on time series $\{x_t\}_{t=1}^{T}$ assumed to be generated by the autoregressive process

$$x_t = \theta x_{t-1} + v_t , \tag{20}$$

where θ denotes the autoregressive parameter and v_t is a white noise process. Suppose the power is based on N_r replications. Let w denote the count of rejections of a false null hypothesis in N_r. Further, let ϑ denote a conventionally accepted number of rejections of a false null hypothesis in the N_r replications required for a good hypothesis testing procedure. Define the statistic

$$Z_\omega = \sum_{i=1}^{m} \kappa(\vartheta) D_i^2(\omega, \vartheta) , \tag{21}$$

where $\kappa(\vartheta) = \vartheta^{-1}$, $D_i^2 = (\omega_i - \vartheta)^2$ and m denotes the range of values of the autoregressive parameter θ considered in the simulation. The value $\sqrt{D_i^2(\omega, \vartheta)}$ represents the deviation of ω from ϑ. It is obvious from (21) that the statistic Z_ω is small if the deviation $\sqrt{D_i^2(\omega, \vartheta)}$ is small. On the other hand, Z_ω is large when $\sqrt{D_i^2(\omega, \vartheta)}$ is large. Thus the magnitude of the deviation of ω_i from ϑ is informative about the size of the statistic. The statistic Z_ω can be employed as a unit root test selection criterion such that different tests can be rated. A practical criteria of choice with the potential for high discriminatory power would be the one based on smaller values of the statistic Z_ω or a function of it, e.g., $S_\omega = \sqrt{Z_\omega}$. There is flexibility with the choice of ϑ in that it can be set by the experimenter in consideration with the objective of the experimenter. The most simple and convenient way to set ϑ is via setting a threshold on the number of replications on which the power is based. That is, we use

$$\vartheta = \delta N_r , \tag{22}$$

where δ represents the threshold and satisfies the condition $0 < \delta < 1$. Higher values of δ, particularly values close to 1 are appropriate since they correspond to the test with high power, using power as a criterion of choice. There are several plausible values of δ which are close to 1 to select from and the question here is how to select the best value. We discuss some methods for choosing appropriate values of ϑ next.

3.1 Selection of Conventional Benchmark Value (ϑ)

The choice of ϑ in (21) is primal to the performance of the proposal since different choice of the threshold δ will yield different result for the most reliable test. The choice of the appropriate ϑ reduces to choosing the best δ. One can perform sensitivity analysis based on all possible values of δ close to 1 to examine the region of stability of the scores Z_ω in order to select the optimal value of ϑ. Note that optimal here implies the most stable value of ϑ that yields most reliable test. This amounts to applying (21) with

$$\delta = \{0.8, 0.805, \ldots, 0.995, 1\}, \tag{23}$$

for the set of δ that yields the same result for most reliable tests. The rationale for the choice of δ above is simply based on the convention that hypothesis test with power at least 0.8 is considered a good test.

An attractive alternative approach to the choice of δ above is based on average score over all possible values of δ specified in (23), leading to scores of the form

$$\bar{Z}_\omega = \frac{1}{K} \sum_{j=1}^{K} \sum_{i=1}^{m} \kappa_j(\vartheta) D_{ij}^2(\omega, \vartheta) , \tag{24}$$

where $\kappa_j(\vartheta) = \vartheta_j^{-1}$, $D_{ij}^2(\omega, \vartheta) = (\omega_i - \vartheta_j)^2$, $\vartheta_j = \delta_j N_r$, K represents the number of values in the set δ given in (23). Notice that (24) is a simple modification of (21) to take into account all possible conventional choices for good unit root hypothesis.

These criteria will be utilized in our experimentation later. These proposals are novel approaches in an attempt to address the selection of most reliable unit root test in time series analysis. The motivation for the above proposals is derived from the Chi-square statistic and its utility in novelty detection (Ye & Chen, 2001). Also, by the central limit theorem, if the number of parameter values is large, that is $m > 30$, then the random variable Z_{ω} approaches the normal distribution. The approximate normality of the scoring statistic allows the development of robust criteria for selection of the most reliable unit root test. It is also possible to build a novelty detector-based unit root selector that has the ability to prune all values of the AR parameter that results in less power yielding hypothesis test based on Z_{ω}, that is centered on the region of high power. However, this particular attractive feature of the scheme is not explored in this paper.

4 Performance Assessment

We outline the performance criterion that we employ for assessing the unit root tests considered using our proposed scheme. First, we consider two lines of simulations; one tailored towards assessment of the proposed scoring approach and the other towards establishing whether the variance of the error term has effect on the choice of unit root tests. Second, because the autoregressive parameter estimate is fixed in terms of the real data (i.e., single value), the assessment based on the real data application will apply the most reliable test realized from the simulation assessment.

4.1 Simulation 1

We report results on the performance of selected unit root tests in simulation studies conducted based on an autoregressive process of order 1 as given in model (20). We consider time series of varied length with $T = 25, 30, 50, 100(50), 350, 500$. The innovation structure is assumed to follow probability model of the form $v_t \sim N(0, \sigma_o^2)$. The rationale for the choice of T as specified above was to investigate the performance of the selected tests on relatively broad range of time series instead of those considered by Fedorova (2016). We set the number of replications N_r to 1000 and $\sigma_0^2 = 1$. For values of the autoregressive parameter, we consider set of values satisfying stationary AR(1) process so that $0 < \theta < 1$. In particular, we utilize set of length 99 of the form $\{0.01, 0.02, 0.03, \ldots, 0.99\}$ for our simulation studies. To avoid randomness associated with generating series of length T independently that could compound the randomness from the innovation (non-systematic errors), we first create a database of size $1000 \times N_r$ simulated time series for each value of θ and then series of length T as specified above are created for each θ. The selected unit root tests reviewed in Section 2 are applied to each time series. The count of rejections of the null hypothesis of non-stationarity among the N_r replications for each of the simulated time series when series are generated from a stationary AR(1) process are monitored for computing the power, $(1-\beta)$ of all the tests, with the exception of the KPSS test. For KPSS test, the count of correct decisions made by the test in the number of replications are monitored and its proportion computed as its performance measure due to the nature of its hypothesis. Also we monitor the rejections of the null hypothesis of stationarity in N_r when the time series are simulated from a non-stationary autoregressive process to assess the power of the test. To ensure completeness of the simulation, we also assess the size of each unit root test with the exception of KPSS test, by simulating from model (20) with $\theta = 1$. All codes were written in R and run on an Intel (R) quinque processor Windows PC 5:2 Ghz workstation.

4.2 Simulation 2

Assessing the impact of Error Structure on Unit Root Test

The goal of this assessment is to investigate the impact of the error structure on the choice of unit root test. This simulation follows the same setting as in Simulation 2 except that the fixed error variance σ_0^2 is assumed variable. In particular, we consider for σ_0^2 values $\{0.01, 0.05, 0.5, 0.8, 0.9, 1, 2\}$. Results are presented in the Appendix.

4.3 Performance of Selected Unit Root Tests based on Power

4.3.1 Dickey Fuller Test

Figure 1 is a graphical representation of the power functions of DF test for simulated time series at different lengths. Dickey Fuller test is characterized by low power for large values of the parameter. For $\theta < 0.3$, DF test exhibits a better performance in terms of power (power around 1). For short time series $(T = 25, 30, 50)$, the power of the test gradually decreases inversely to the value of θ. For $\theta > 0.5$, the test is more likely to not reject the non-stationarity of the time series despite the series actually being stationary. If the time series has a medium length $(T = 100, 150, 200)$, the power increases until it reaches $\theta > 0.85$ where it begins to fall, suggesting the test inability to distinguish between stationarity and non-stationarity for values of θ close to one (1). It is clear that Dickey Fuller test becomes very reliable in the case where time series has a large length $(T = 250, 300, 500)$, regardless of the parameter values. It can be inferred from the power functions that DF test could be reasonably good for very large series, a feature that will be investigated latter.

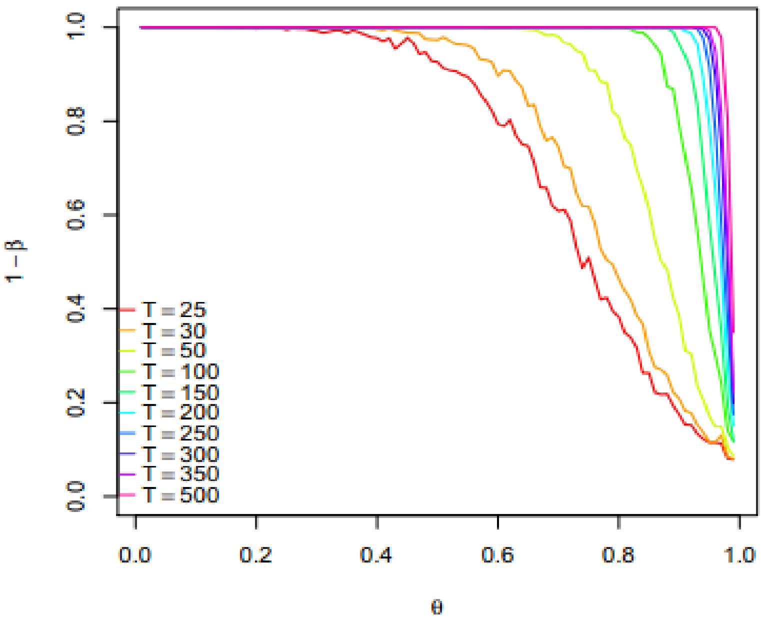


Figure 1: Power function of DF test for simulated time series of various lengths.

4.3.2 Augmented Dickey Fuller

The typical behaviour of ADF test in terms of the power function for time series of varying lengths is illustrated in Figure 2. For short series $(T = 25, 30, 50)$, though the power function decreases steadily over the range of θ the difference in the power functions can easily be distinguished. The test properties improve for time series of medium length. It can be observed that the test's performance improves slightly for time series of medium length $(T = 100, 150, 200)$. The power of the test remains high until the value $\theta = 0.8$ and begins to decrease. Clearly, the test exhibits better performance in long series $(T = 250, 300, 500)$, for values of θ between 0 and 0.97.

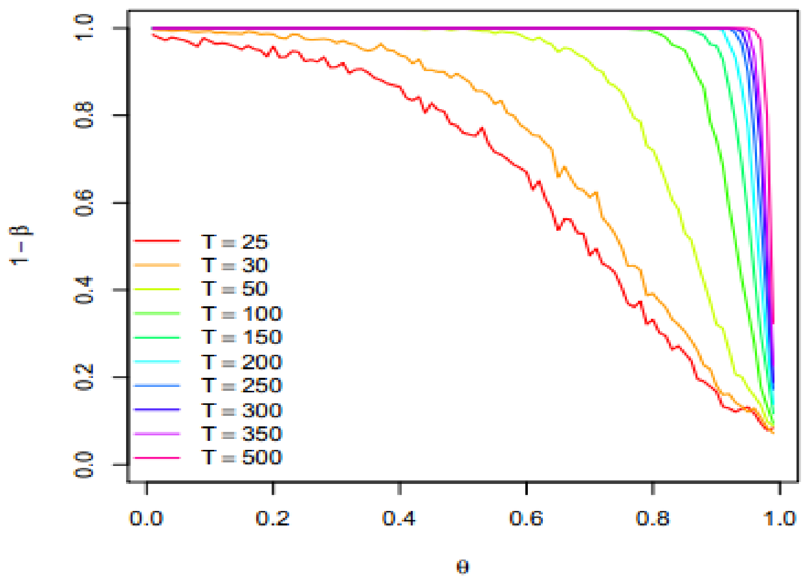


Figure 2: Power function of ADF test for simulated time series of various lengths.

4.3.3 Phillip Perron Test

The power functions of the unit root test based on a non-parametric adjusted test (Phillip Perron test) is shown in Figure 3. For short time series $(T = 25, 30, 50)$, the power of the test is low as the value of the parameter increases. For values $\theta > 0.2$ the test is more likely to not reject the non-stationarity of the time series despite the fact that it is stationary. The test properties improve for time series of medium length. We observe that for time series of medium length $(T = 100, 150, 200)$, the power is significantly high. The power of the test increases until it reaches value $\theta > 0.8$ and then falls. However, the test achieves maximum success of almost 100% for $T = 250, 300, 500$ for values of θ less or equal to 0.97, thus suggesting the potential to distinguish between stationarity and non-stationarity when even the value of θ is near unity especially for relatively long series.

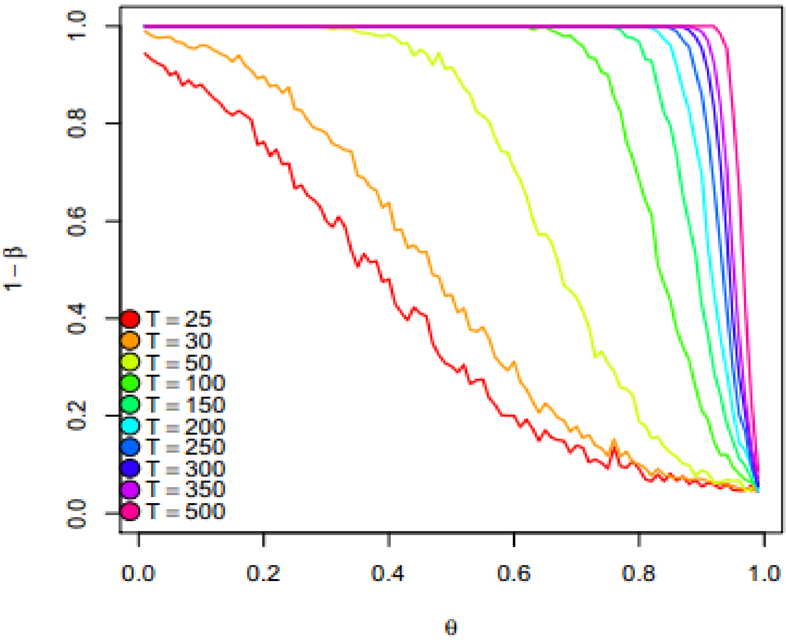


Figure 3: Power function of PP test for simulated time series of various lengths.

4.3.4 Augmented Dickey Fuller-Generalized Least Square Test

Figure 4 gives the power functions for ADF-GLS test for simulated time series of varying lengths considered for the study. Compared to all the above power functions, ADF-GLS test exhibits noticeable performance for medium to large series $(T = 100, 150, 200, 250, 350, 500)$ with the power being at least 80%. However, for short series $(T = 25, 30, 50)$, the test is modest and has high tendency not to reject the null hypothesis when it is false. In general, the power characteristics of ADF-GLS test for values of close to one are fairly smoother compared to those discussed above especially for medium to long time series.

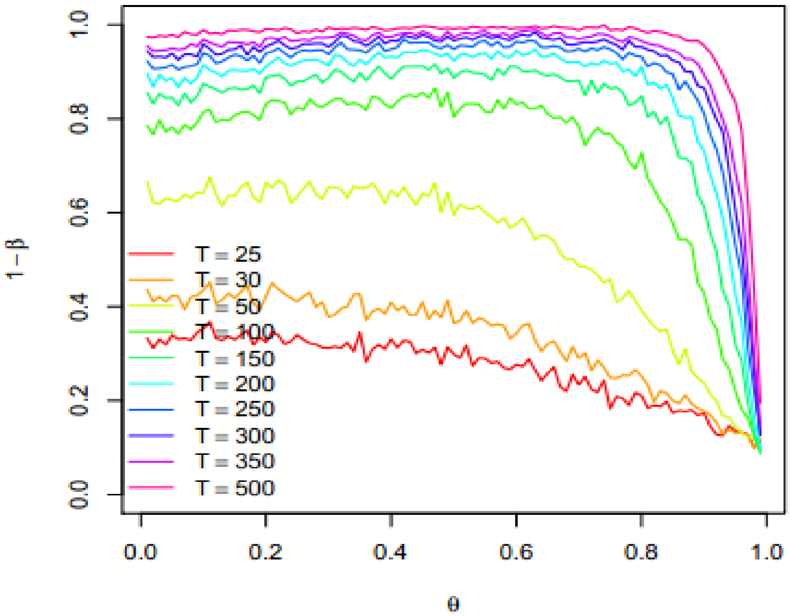


Figure 4: Power function of ADF-GLS test for simulated time series of various lengths.

4.3.5 *Schmidt and Phillip Test*

The nature of the power functions of Schmidt and Phillips (SP) test for simulated time series considered for the study at different lengths is shown in Figure 5.

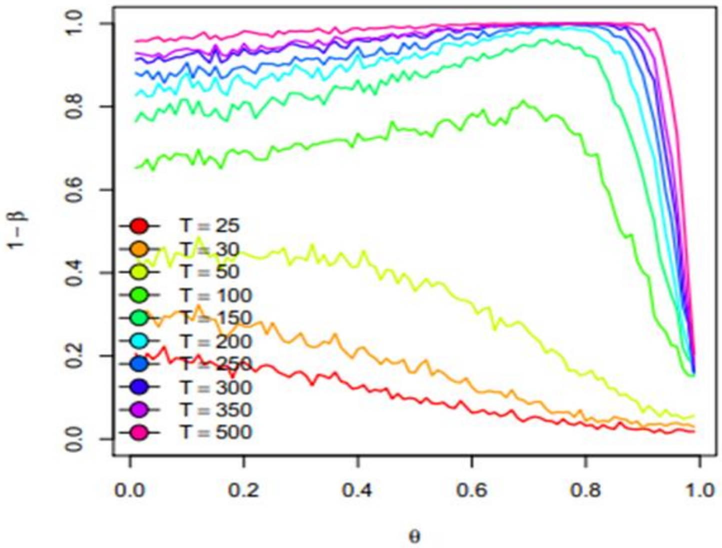


Figure 5: Power function of SP test for simulated time series of various lengths.

The power functions for medium to long series increases gradually with increasing values of the parameter θ values except for relatively large values close to one for which it exhibits some smooth decline. This however indicates the potential of SP to distinguish between stationarity and non-stationarity for the above mentioned time series. Nevertheless, SP test clearly demonstrates the challenge it has for series of relatively short length regardless of plausible values of θ. This is evident in the power function graph with a relatively wide separation between the various time series considered.

4.4 Comparison of Unit Root Tests Based on Classification of Series using Length

This section compares all the relevant unit root tests considered according to their ability to determine the presence of unit root or stationarity for different groups of time series based on length. In this assessment, we empirically put the simulated time series into three main groups; Short time series: $T = \{25, 30, 50\}$, Medium time series: $T = \{100, 150, 200\}$, and Large time series: $T = \{250, 300, 350, 500\}$. We then select the third member of each group for each test for comparison.

4.4.1 Short Time Series

Figure 6 is the power functions for the selected unit root tests for relatively short time series. It is evident from the power functions that Schmidt and Phillips test and Augmented Dickey-Fuller Generalized Least Square test lack the ability to distinguish between the presence of unit root or otherwise. Dickey Fuller test, Augmented Dickey Fuller test and Phillip Perron test record a power value of one for $\theta < 0.3$. Dickey Fuller test and Augmented Dickey Fuller test are suitable in this situation.

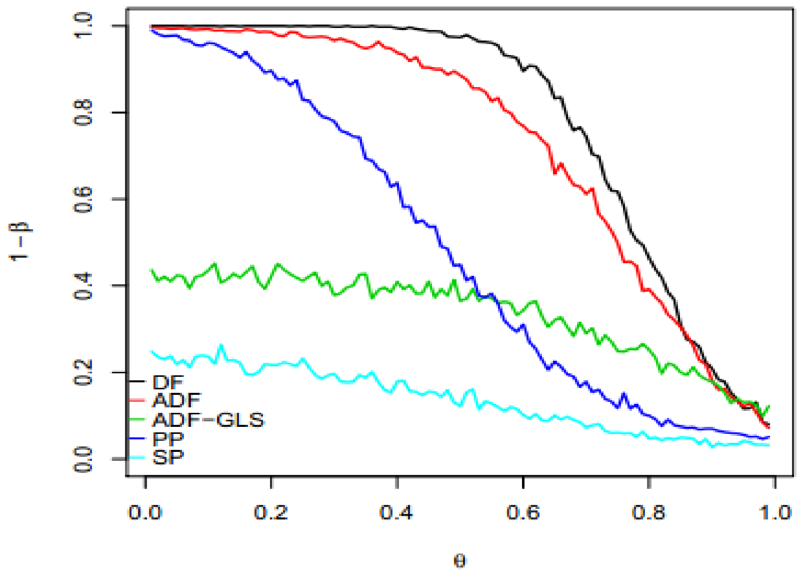


Figure 6: Comparison of power functions of the unit root tests for time series of short length.

4.4.2 Medium Time Series

The dynamics of power functions for the selected time series considered on a group of time series classified as medium in the simulation study is presented in Figure 7. Although all the tests showed improved performance, it is evident that DF, ADF, and PP tests out-perform the other unit root tests for assessing stationarity and non-stationarity. However, it can be observed that the power of all the tests fluctuates within 0.8 and 1 for $\theta < 0.8$. Afterwards, it sharply declines for values of θ approaching 1. Furthermore, it can be observed that their powers exhibit slight fluctuation around one for $\theta < 0.9$.

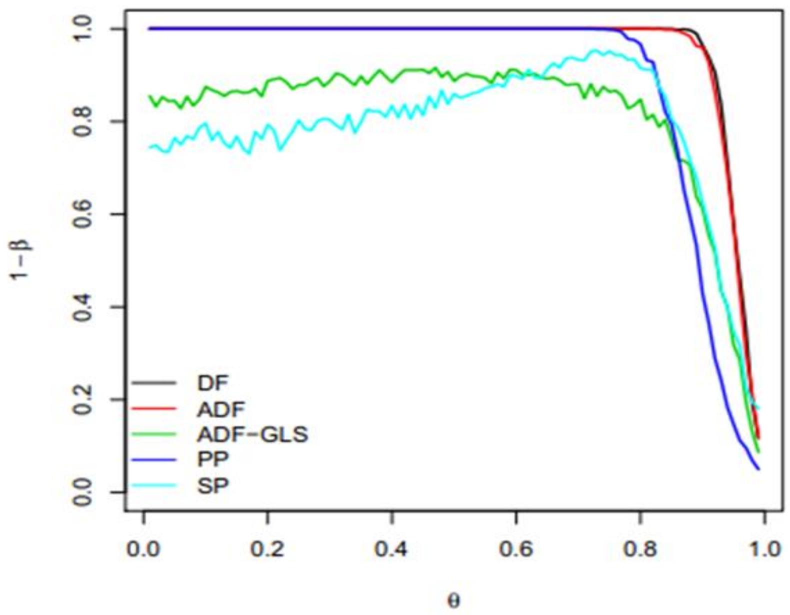


Figure 7: Comparison of power functions of the unit root tests for time series of medium length.

4.4.3 Large Time Series

The characteristics of power functions for the selected time series considered for long series in the simulation studies is presented in Figure 8. An obvious observation from the graph is that all the tests are showing good performance with Dickey-Fuller, Augmented Dickey-Fuller and Phillip-Perron tests being almost indistinguishable. Also, it is evident that the power of the three exhibit, slight fluctuation around one (1) for all values of θ less than 0.9.

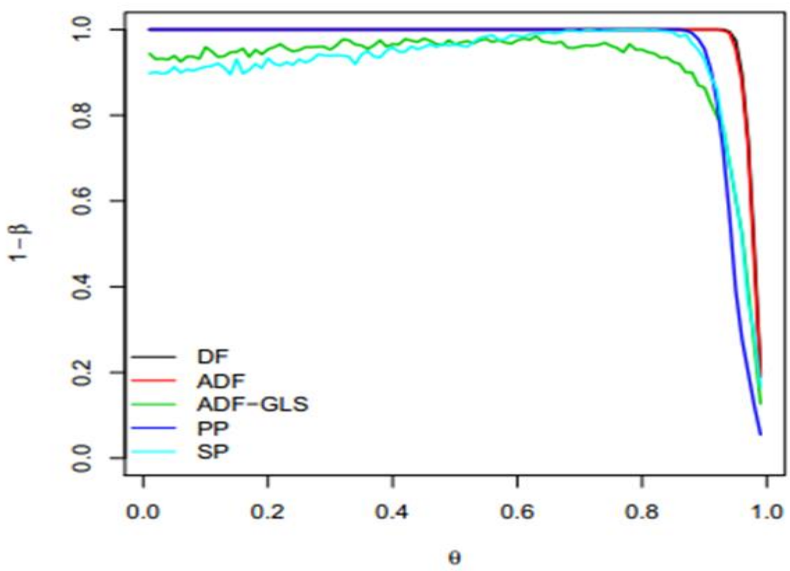


Figure 8: Comparison of power functions of the relevant tests for time series of large length.

The challenge associated with the use of power based approach for selecting the most reliable unit root test in time series analysis is illustrated in the above results.

Table 1: Size of selected tests for simulated time series of length T at $\alpha = 0.005$

Length (*T*)			Test		
	DF	ADF	ADF-GLS	PP	SP
25	0.06	0.07	0.09	0.05	0.04
30	0.06	0.06	0.09	0.05	0.04
50	0.05	0.05	0.10	0.06	0.07
100	0.04	0.04	0.08	0.06	0.16
150	0.04	0.06	0.06	0.07	0.18
200	0.05	0.07	0.07	0.08	0.15
250	0.05	0.06	0.07	0.07	0.14
300	0.05	0.06	0.06	0.07	0.13
500	0.04	0.05	0.06	0.06	0.10

Information on the size of the various unit root tests considered for simulated time series of varied length at standard level of significance $(\alpha = 0.05)$ is presented in Table 1. The size of a hypothesis test is the probability of type I error, α .

4.5 Performance of Selected Unit Root Tests Based on Scoring Scheme

4.5.1 Sensitivity Analysis of the Score (Z_ω)

The performance of the non-averaging scoring proposal for a range of conventional benchmark (i.e., δ) is shown in Figure 9. It can be seen that the scheme exhibits some degree of sensitivity to the value of δ , especially when $\delta = 0.8$, but stabilizes for $\delta \geq 0.9$. Thus, for the application of the scores Z_ω , for deciding on the most reliable unit root test, it is appropriate to set the benchmark δ as $\delta \geq 0.9$ for stable performance. Furthermore, the graph shows that optimum length of series for which the best test based on the score Z_ω can be detected is about $T = 150$ .The basis for the performance comparison based on this scheme is simply the smaller the score Z_ω , the better the test. For short time series, the DF and ADF are consistently associated with small Z_ω scores for all values of δ .

ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.8, No.7, 2018

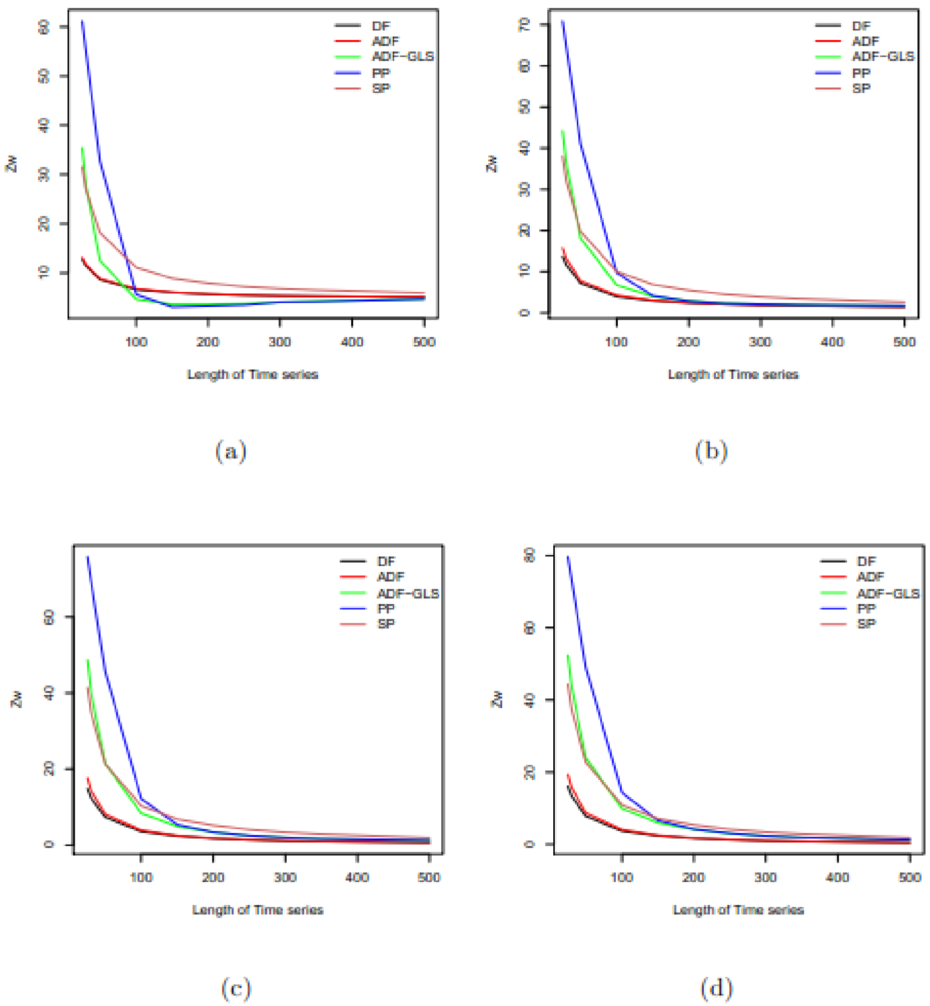


Figure 9: Comparison of score functions of the relevant tests for varied δ based on time series of varied length: (a) $\delta = 0.8$ (b) $\delta = 0.9$ (c) $\delta = 0.95$ and (d) $\delta = 0.99$.

Average Scores

Figure 10 shows the nature of the selected unit root tests under the proposed scoring approach based on the average scoring scheme, over a range of time series lengths considered for choosing between competing tests. In general, arranging the test based on the scores yields DF, ADF, PP, ADF-GLS, and SP test. Considering regions of low scores across the tests corresponding to series of appreciable length $(T > 100)$ facilitates the choice of the best unit root test regardless of the proximity of the parameter to one (1). Thus, it is evident that the DF and ADF tests are the most reliable tests for the unit root of time series.

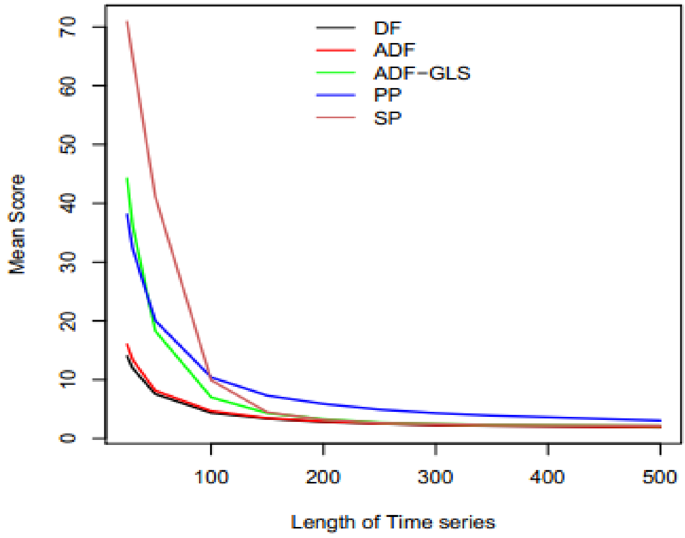


Figure 10: Similarity-based scoring scheme results for unit root tests.

4.5.2 Confirmatory Test Based on KPSS Test

The KPSS test tests the null hypothesis of stationarity against the alternative hypothesis of non-stationarity. As a result, we monitor the proportion of correct decisions in testing the above hypothesis when indeed the null hypothesis is true. Also, the power of the test when θ is set to one (1) is computed as the proportion of rejection of the null hypothesis out of the number of replications. The nature of KPSS test in terms of making correct decisions over a range of autoregressive parameter values is shown for time series of varied lengths in Figure 11. It is evident from the plot that the KPSS test has the ability to make the correct decision regardless of the length of the series. Figure 12 gives the typical nature of KPSS test in terms of power for autoregressive parameter value for time series of varied length considered in the simulation studies. The strength of KPSS test is significantly evident. Now combining the information provided in the $(1-\alpha)$ and $(1-\beta)$ functions, KPSS test exhibits a potential to be a rich source of benchmark for selection of best unit root test. As a result, we propose that KPSS test be utilized as a confirmatory test for the other unit root tests.

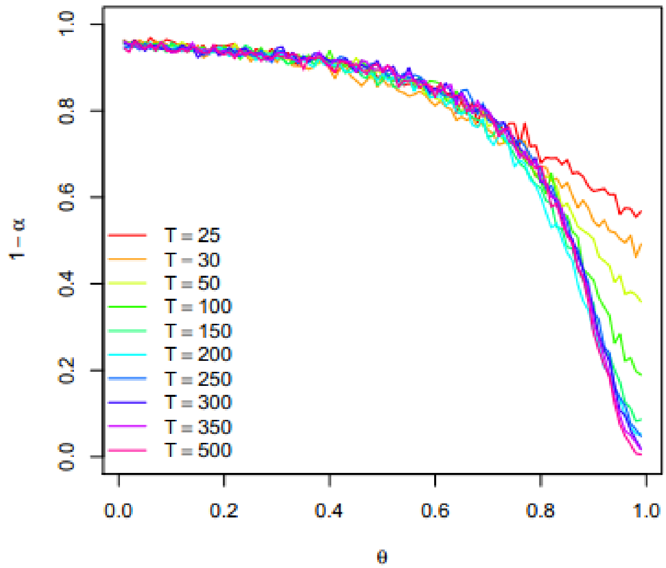


Figure 11: The proportion of correct decisions $(1-\alpha)$ over a range of autoregressive parameter θ, of KPSS test for simulated time series

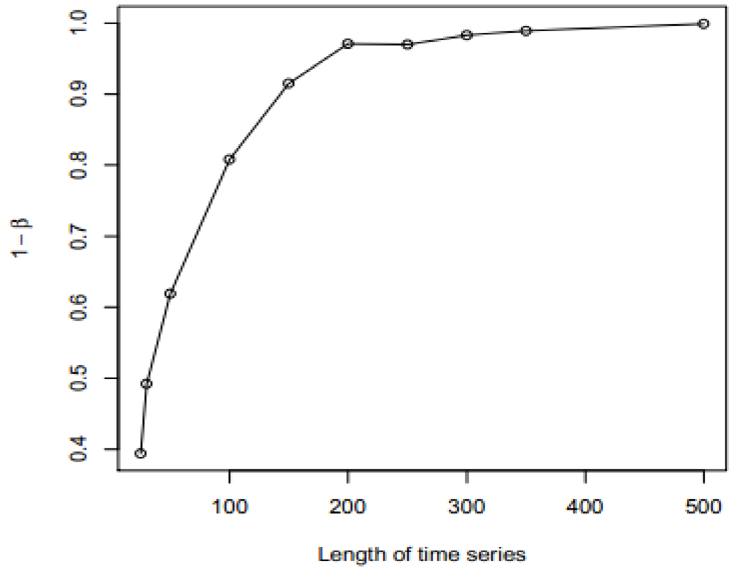


Figure 12: The power function of KPSS test for simulated time series of varied length (T) based on autoregressive parameter $\theta = 1$.

4.6 Real Data Application

Application of the two most reliable unit root tests obtained from the simulation studies to real time series of University of Cape Coast halls electricity consumption data and time series of some macro-economic variables is considered in this section. First, we give a brief description of the data before examining the results. The electricity consumption dataset is a 132×5 matrix of recent ten (10) years monthly observations involving five out of six halls of residence of the University of Cape Coast. The sixth hall was deleted due to the presence of missing values. The electricity consumption data is obtained from the electricity section of the University. The application here is based on the results of the differenced time series which are found to be stationary. Results are presented in Table 2, and Table 3 where the following notations have been utilized.

x_1 : Interest rate, x_2 : Exchange rate, x_3 : Inflation rate on food,

x_4 : Inflation rate on non-food, x_5 : Monetary policy rate.

Data on these macro-economic variables also cover the same 10 years period and were obtained from the website of the Bank of Ghana (w.w.w.bog.gov.gh).

Table 2: Application of Dickey-Fuller test on first differenced Datasets at $\alpha = 0.05$.

University Halls Electricity Consumption Data			Macro-Economic Data		
Hall	Test Statistic	Table value	Variable	Test Statistic	Table Value
1	−10.939	−1.95	x_1	−10.330	−1.95
2	−13.809	−1.95	x_2	−12.785	−1.95
3	−15.661	−1.95	x_3	−11.002	−1.95
4	−10.260	−1.95	x_4	−17.441	−1.95
5	−11.255	−1.95	x_5	−10.668	−1.95

Table 3: Application of Augmented Dickey-Fuller (ADF) test on first differenced data at $\alpha = 0.05$.

University Halls Electricity Consumption Data			Macro-Economic data		
Hall	Test Statistic	Table value	Variable	Test Statistic	Table Value
1	–10.940	–1.95	x_1	–7.081	–1.95
2	–10.653	–1.95	x_2	–6.392	–1.95
3	–10.019	–1.95	x_3	–7.529	–1.95
4	–9.935	–1.95	x_4	–12.638	–1.95
5	–10.493	–1.95	x_5	–5.543	–1.95

Table 2 shows the results of application of DF test on the first differenced time series data obtained on electricity consumption and macro-economic variables considered.

The test exhibits test statistic values greater in absolute values than the table values for all five halls of residence, implying that the series does not contain a unit root. This is also true for all macro-economic variables. In Table 3, similar results are obtained for the ADF test for both datasets.

5 Conclusion

We have proposed and implemented a simple and easy to apply similarity-based schemes based on information underpinning power in hypothesis testing for identifying the most reliable unit root tests in time series analysis. Application of the proposals to simulated time series processes of varying length show their utility. Specifically, Dickey-Fuller (DF) and Augmented Dickey-Fuller (ADF) Tests were the two most reliable unit root tests. The most reliable tests from the simulation studies are applied to real time series data of University of Cape Coast residential Hall electricity consumption and time series of some macro-economic variables. Thus, an appropriate criteria have been developed for unit root tests selection especially when prior knowledge of the autoregressive parameter value is not available. In this case, the experimenter is able to avoid or reduce the time and resources expended on analyzing the power function point-wise based on the length of the series. Furthermore, we have explored the potential of KPSS test as a confirmatory test for assessing the other available unit root tests thus a plausible building block for developing robust unit root test selection criterion of choice.

Acknowledgement

This research was fully supported by Samuel and Emelia Brew-Butler-SGS/GRASAG research grant for which we are very grateful.

References

Box, G. & Jenkins, G. (1970). Time series analysis-forecasting and control. san francisco: Holden day. 553

Dickey, D.A., & Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association,* 74(366a), 427-431.

Dickey, D.A. (1976). Estimation and hypothesis testing in nonstationary time series. PhD
Thesis, Iowa State University, 1976. https://lib.dr.iastate.edu/rtd/6267

Dickey, D.A. & Fuller, W.A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society,* 1057-1072.

Elliott, G.., Rothenberg, T.J., & Stock, J.H. (1996). Efficient tests for an autoregressive unit root. Econometrics

64 (4): 813-836. *MATH CrossRef MathSciNet Google Scholar*.

Fedorova. D. (2016). Selection of unit root test on the basis of length of the time series and value of AR(1) parameter. *STATISTIKA*, 96(3):3.

Fuller, W.A. (1976). *Introduction to statistical time series*, New York: John Wiley.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1-3):159-178.

Ng, S., & Perron, P. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association*, 90 (429), 268-281

Phillips, P.C.B. & Perron. P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2):335-346.

Schmidt, P., & Phillips, P.C.B. (1992). Lm tests for a unit root in the presence of deterministic trends. *Oxford Bulletin of Economics and Statistics*, 54(3): 257-287.

Schwert, G..W. (2002). Tests for unit roots: A monte carlo investigation. *Journal of Business & Economic Statistics*, 20(1), 5-17.

Ye, N. & Chen, Q. (2001). An anomaly detection technique based on a chi-square statistic for detecting intrusions into information systems. *Quality and Reliability Engineering International*, 17(2):105-112.

Zapata, H.O., Maradiaga, D.I., Pujula, A.L., & Dicks, M.R. (2011). Recent developments in unit root tests and historical crop yields. In *Selected Paper Prepared for Presentation at the Agricultural and Applied Economics Association's 2011 AAEA & NAREA Joint Annual Meeting, Pittsburgh, Pennysylvania, 24-26*.

Appendix

Results: Assessment of the impact of Error Structure on Unit Root Test

The rest of the results of Simulation 2 in Section 4.2 in which the variance of the innovation term was varied over some selected range are shown in Figure A1 and Figure A2. Apparently, it is evident that the choice of error variance of an AR process exhibits no impact on the power of a given unit root test. Thus, the error variance is not informative about the choice of the most reliable unit root test.

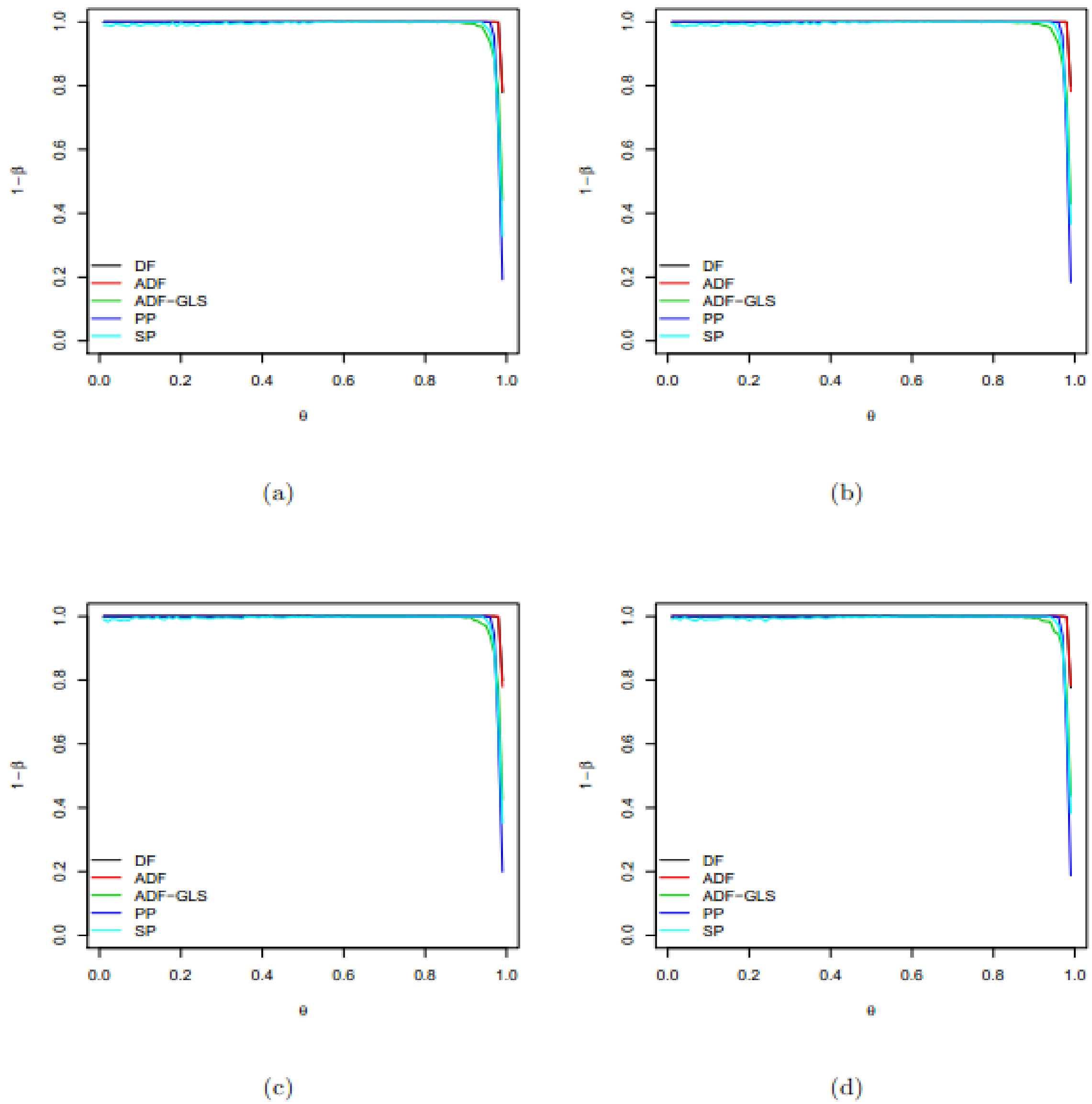
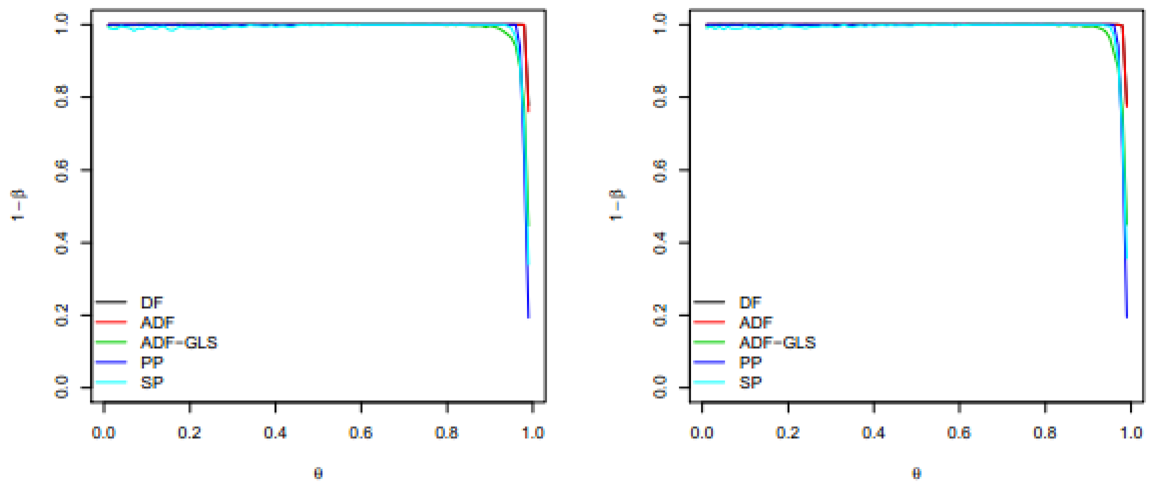


Figure A1: Comparison of power functions of the unit root tests for time series of very long length with varied error variance: (a) $\sigma^2 = 0.01$ (b) $\sigma^2 = 0.05$ (c) $\sigma^2 = 0.5$ and (d) $\sigma^2 = 0.8$.



(a) (b)

Figure A2: Comparison of power functions of the unit root tests for time series of very long length with varied error variance: (a) $\sigma^2 = 1$ and (b) $\sigma^2 = 2$.

Assessing the Likelihood for Equal Performance of Selected Unit Root Tests

We also considered another simulation where the length was extended to 1000 in order to explore further the above observation. The result is presented in Figure A3. The plot is the power functions of the unit root tests considered in the study for time series of very large length, in this case $T = 1000$. It is apparent that the tests are almost indistinguishable adding value to their potential for better performance for long series rather than short series regardless of the proximity of the parameter value to 1. Generally, based on the assumption under which the various tests were constructed together with their dependence on the length of the series as we have seen in our experimentation, the tests would be appropriate for assessing the presence of unit root in a relatively long series.

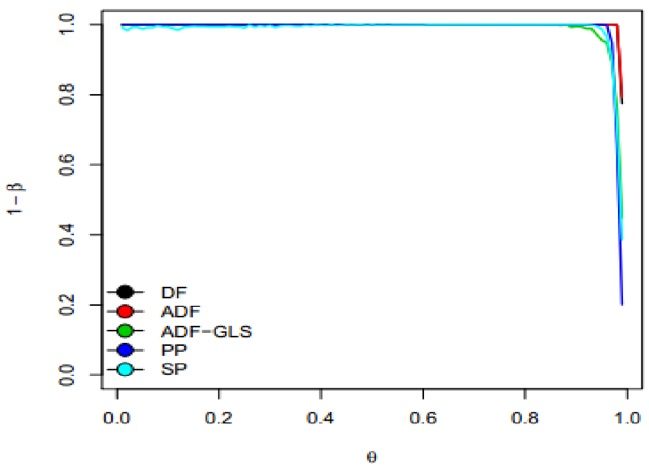


Figure A3: Power functions of selected unit root tests for time series of length $T = 1000$.