

Evaluating the Performance of Unit Root Tests in Single Time Series Processes

Jonathan Kwaku Afriyie¹, Sampson Twumasi-Ankrah^{1,*}, Kwasi Baah Gyamfi², Doris Arthur¹,
Wilhemina Adoma Pels¹

¹Department of Statistics and Actuarial Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

²Department of Mathematics, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Received August 12, 2020; Revised October 9, 2020; Accepted October 19, 2020

Cite This Paper in the following Citation Styles

(a): [1] Jonathan Kwaku Afriyie, Sampson Twumasi-Ankrah, Kwasi Baah Gyamfi, Doris Arthur, Wilhemina Adoma Pels, "Evaluating the Performance of Unit Root Tests in Single Time Series Processes," *Mathematics and Statistics*, Vol. 8, No. 6, pp. 656 - 664, 2020. DOI: 10.13189/ms.2020.080605.

(b): Jonathan Kwaku Afriyie, Sampson Twumasi-Ankrah, Kwasi Baah Gyamfi, Doris Arthur, Wilhemina Adoma Pels (2020). *Evaluating the Performance of Unit Root Tests in Single Time Series Processes*. *Mathematics and Statistics*, 8(6), 656 - 664. DOI: 10.13189/ms.2020.080605.

Copyright©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract Unit root tests for stationarity have relevancy in almost every practical time series analysis. Deciding on which unit root test to use is a topic of active interest. In this study, we compare the performance of the three commonly used unit root tests (i.e., Augmented Dickey-Fuller (ADF), Phillips-Perron (PP), and Kwiatkowski Phillips Schmidt and Shin (KPSS)) in time series. Based on literature, these unit root tests sometimes disagree in selecting the appropriate order of integration for a given series. Therefore, the decision to use a unit root test relies essentially on the judgment of the researcher. Suppose we wish to annul the subjective decision. In that case, we have to locate an objective basis that unmistakably characterizes which test is the most appropriate for a particular time series type. Thus, this study seeks to unravel this problem by providing a guide on which unit root tests to utilize when there is a disagreement between them. A simulation study of eight (8) univariate time series models with eight (8) different sample sizes, three (3) differencing orders, and nine different parameter values were performed. It was observed from the results that the performance of the three tests improved as the sample size increased. Based on comparing the overall performance, the KPSS was the "best" unit root test to use when there is disagreement.

Keywords Augmented Dickey-Fuller, KPSS, Phillips-Perron, Simulation, Time Series Analysis, Unit Root Tests

1. Introduction

A time series is a sequence of ordered data from a family of random variables, say $\dots, x_{t-1}, x_t, x_{t+1}$ Tebbs [1]. Time series modeling aims to study past observations of a series to develop an appropriate model that explains the series Adhikari [2]. A time series process is said to be weak stationary if its mean, variance, and covariance do not change over time Test [3]. A time series is characterized as integrated of order (d) if their stationarity is achieved by differencing the series "d times" Fedorov'a [4]

The importance of stationarity cannot be underestimated since many statistical tests and forecasts in time series depend on it. It is necessary to ensure that a time series data is stationary before building a model. Non-stationary time series exhibits trends, seasonal variations from which these models' forecasts are not reliable Twumasi-Ankrah [5].

Therefore, it is vital to make a non-stationary time series data stationary by removing the time trends and/or seasonal variations before carrying out any time series analysis. According to Brownlee[6], stationarity can be measured in the following ways: through time plots (i.e. plotting the data and visually checking if there is any obvious trends or seasonality); summary statistics (i.e. reviewing the mean and variance for your data for seasons or random partitions and checking for significant difference) and statistical tests (i.e. using statistical tests to check if the expectations of stationarity are met or have been violated).

There are two general approaches for testing stationarity, namely parametric and non-parametric Masliah [7]. This study focuses on the parametric that is the unit root test. A unit root is a characteristic of a stochastic process that can create problems in statistical analysis. A stochastic process is said to have a unit root if (1) is a solution to the process's characteristic equation. In this case, the process is said to be non-stationary. Several statistical tests have been employed to differentiate if a series has a stochastic trend and should be differenced or if it has a deterministic trend and should be detrended Cochrane [8]. Non-stationarity in mean can be a deterministic trend or stochastic trend. In deterministic trend, the series is a linear function of time, that is $x_t = \alpha + \beta t + \varepsilon_t$. The parameter β measures the average change in x_t from one period to the other. According to Mushtaq [9], macroeconomic data has stochastic trends, and therefore, such series has a unit root, and using these variables in modeling may lead to spurious regression. That is, regressing on trended series is meaningless and yields misleading outcomes.

Determining and verifying the order of integration is a very wide area that incorporates broad rundown of tests described as unit root tests where some of them are, Dickey and Fuller (DF) test, Augmented Dickey-Fuller (ADF) test, Phillips-Perron (PP) test, Kwiatkowski Phillips Schmidt and Shin (KPSS) test, ADF-GLS test, Ng-Perron, etc. [4]. In literature, the three most commonly used tests are the ADF test, the PP test, and the KPSS test [4].

Davidson [10] reports that the PP test performs worse in finite samples than the ADF test. Again, according to Schwert [11], tests for unit roots can have low power against some particular other options. From our preliminary analysis, these tests disagree in selecting the appropriate order of integration for a given series.

In our simulation study, we are interested in investigating three main aspects, namely: (1) performance of the unit root tests by varying the order of integration, (2) performance of the unit root test by varying sample size, and (3) the overall performance of the three tests used in this study.

2. Materials and Methods

This section seeks to outline the specific procedures or techniques used in this work. It includes how the simulation was done to examine the three commonly used unit root tests' performance: ADF, PP and KPSS.

2.1. Review on the Unit Root Tests

The Augmented Dickey-Fuller (ADF) Test

The ADF test includes extra lagged in terms of the dependent variables to dispense with autocorrelation. One of the weaknesses is that the power of the test is very low. The ADF test is based on an autoregressive process with order one, which is given by the equation below:

$$x_t = \beta_1 x_{t-1} + \varepsilon_t \quad ; \quad t = 1, \dots, T \tag{1}$$

where β_1 is the parameter for the autoregressive model.

Hypothesis

$$H_0: \hat{\beta}_1 = 1$$

$$H_1: \hat{\beta}_1 < 1$$

Test Statistics:

$$T_{DF} = \frac{\hat{\beta}_1 - 1}{s_{\hat{\beta}_1}} \tag{2}$$

The Phillips Perron (PP) Test

The PP test builds on the DF test. It makes a non-parametric correction to the t-test statistic. One advantage of the PP test over the ADF test is that the PP test is robust to general forms of heteroscedasticity in the error term ε_t . Also, the user does not have to define a lag length for the test regression.

Hypothesis:

H_0 : The process has a unit root.

H_1 : The process does not have a unit root.

Test statistic:

$$Z_\beta = (\hat{\beta}_T - 1) - \frac{1}{2} \frac{T^2 S_\beta^2}{S_T^2} (S_{LT}^2 - S_T^2) \tag{3}$$

$$Z_T = \left(\frac{S_T}{S_{LT}} \right) T_{DF} - \frac{1}{2} (S_{LT}^2 - S_T^2) \frac{1}{S_{LT}} \frac{TS_\beta}{S_T} \tag{4}$$

where

$$T_{DF} = \frac{\hat{\beta}_1 - 1}{S_\beta}$$

$$S_T^2 = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^2$$

$$S_{LT}^2 = S_T^2 + 2 \sum_{j=1}^q (1 - \frac{j}{q+1}) \hat{\gamma}_{j,T} \text{ and}$$

$$\hat{\gamma}_{j,T} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-j}$$

The Kwiatkowski Phillips Schmidt and Shin (KPSS) Test

KPSS test is broadly utilized in empirical work as an accompaniment to standard unit root tests Hadri[12], and it is not affected by seasonal dummies Phillips[13].

Hypothesis:

H_0 : The series is stationary

H_1 : The series is not stationary.

Test statistic:

$$LM = \sum_{t=1}^T \left(\frac{S_t^2}{\hat{\sigma}_\varepsilon^2} \right) \quad (5)$$

Where

$S_t = \sum_{t=1}^T \varepsilon_t$ and $\hat{\sigma}_\varepsilon^2$ is the estimate of variance σ_ε^2 of process ε_t from the equation.

2.1. Simulation

The simulation was done using existing codes in R version 3.1.1.

The algorithm for the simulation is given below:

- Generate data from an AR process of order 1 to 4 and differencing order from 0 to 2 (i. e. ARI(p, d), p = 1, 2, 3, 4, d = 0, 1, 2) and that of Moving Average (MA) processes from order 1 to 4 and differencing order from 0 to 2 (i.e. IMA (d, q), p=1, 2, 3, 4, d=0, 1, 2).
- For each data generated, positive parameter values (i.e. 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9) were used.
- In the simulation, sample sizes 10, 15, 20, 25, 45, 100, 1000, 10000 were considered.
- For ARI(1,d) and IMA(d,1) processes, nine replications were generated.
- If the sum of parameter values of a model is greater or equal to 1 in ARI(2,d), ARI(3,d), and ARI(4,d), then such models were not considered.
- For IMA(d,2), IMA(d,3), and IMA(d,4), 81, 81, and 207 replications were generated, respectively.
- Perform ADF, PP, and KPSS tests on the model.
- Record the number of times (frequency) in percentage terms; each test selects the true model that generated the data.

Evaluation Procedures

Here, we show the procedures for evaluating the performance of the unit root tests.

Performance of the individual tests in selecting the appropriate order of integration.

The algorithm used is summarized below:

- Count the number of times each model at different sample sizes, and parameter values are stationary.
- Convert the counts to percentages using the formula below:

$$PERC = \frac{n}{S} \quad (6)$$

where,

n is the number of times the data was stationary, and S is the number of replications.

Performance of the unit root tests as model order increases

The algorithm used is summarized below:

- All the AR processes from order 1 to 4 and MA processes from order 1 to 4 with the same integration order were grouped.

- The eight different sample sizes were categorized into two; small and large. For small sample sizes 10, 15, 20, 25 were used (i.e. 10-25) and large sample sizes 45, 100, 1000, 10000 were also used (i.e. 45-10000).
- The average of ARI(p, d), and IMA(d, q) where p, q = 1, 2, 3, 4 and d = 0, 1, 2 which fell into the small sample size category were computed.
- Also the average of AR(p, d) and IMA(d, q) where p, q = 1, 2, 3, 4 and d = 0, 1, 2 which fell into the large sample size category were computed.
- The average percentage values indicating stationarity of each of the model order is computed as

$$AVRG_0 = \frac{1}{m} \sum_{i=0}^m PERC_i \quad (7)$$

where $PERC_i$ are the percentage values, and m is the number of elements in each model order category, which is 4.

Performance of the unit root tests as the sample size increases

In this subsection, the following algorithm was used:

- All the Autoregressive processes with the same order of integration were grouped. i.e.
 - [ARI(1, 0); ARI(2, 0); ARI(3, 0); ARI(4, 0)],
 - [ARI(1, 1); ARI(2, 1); ARI(3, 1); ARI(4, 1)],
 - [ARI(1, 2); ARI(2, 2); ARI(3, 2); ARI(4, 2)]
- The average percentage values were calculated across the model order for each category of the different sample sizes.
- All the Moving Average processes with the same order of integration were grouped. That is,
 - IMA(0, 1); IMA(0, 2); IMA(0, 3); IMA(0, 4),
 - IMA(1, 1); IMA(1, 2); IMA(1, 3); IMA(1, 4),
 - IMA(2, 1); IMA(2, 2); IMA(2, 3); IMA(2, 4)
- The average percentage values were calculated across the model order for each category of the 8 different sample sizes.
- The average percentage values for each category of the sample size is formulated as

$$AVRG_N = \frac{1}{t} \sum_{i=0}^t PERC_N \quad (8)$$

where N is the sample size, and t is the number of elements in each category, 4.

The overall performance of the unit root tests

To determine the overall performance of the three commonly used unit root tests, the following algorithm was used:

- The performance for the respective Unit Root Tests was grouped into performance classes or groups (PG).
- The percentage performance class interval used was formulated as

$$PGI = \frac{Max-Min}{Number\ of\ Classes} \quad (9)$$

where Max and Min are maximum and minimum Percentage Performances, respectively. Here, the maximum value is 100, the minimum is 0, and the number of classes is 5.

$$PGI = \frac{100 - 0}{5}$$

$$PGI = 20$$

The Performance Classes (PG) are listed below:

0 – 20, 21 – 41, 42 – 62, 63 – 83, 84 – 100.

- Assign weights to the Performance Classes from 0, 1, 2, 3, 4, respectively.
- Summation of the product of weight and the frequencies of the performance of the Unit Root Tests is formulated below:

$$Score = \sum \omega_i r_i \tag{10}$$

Where ω_i are the weights, r_i are the frequencies of the

performance classes.

- Rank Scores with the highest score being the best Unit Root Test

3. Results

In this section, the results from the simulation study are presented in terms of the performance of the individual tests in selecting the appropriate order of integration, the effect of the order of the model on the three conventional unit root tests, the impact of sample size on the three unit root tests, and the overall performance of the unit root tests.

3.1. Performance of the Individual Unit Root Test in Selecting Appropriate Order of Integration

Here, we consider the performance of the individual unit root tests.

Table 1. Percentage of selecting stationarity using ADF

Model	Sample Size							
	10	15	20	25	45	100	1000	10000
d=0								
ARI(1,0,0)	22	11	11	0	33	89	100	100
ARI(2,0,0)	17	3	0	11	31	64	100	100
ARI(3,0,0)	11	4	0	4	29	36	100	100
ARI(4,0,0)	24	12	18	24	14	20	100	100
IMA(0,0,1)	0	0	100	100	100	100	100	100
IMA(0,0,2)	11	0	0	7	47	100	100	100
IMA(0,0,3)	42	0	0	0	28	100	100	100
IMA(0,0,4)	0	15	0	0	9	93	100	100
d=1								
ARI(1,1,0)	22	11	11	0	33	89	100	100
ARI(2,1,0)	14	3	0	17	31	64	100	100
ARI(3,1,0)	11	4	0	7	29	46	100	100
ARI(4,1,0)	24	12	18	24	14	20	100	100
IMA(0,1,1)	0	0	100	100	100	100	100	100
IMA(0,1,2)	10	0	0	0	47	100	100	100
IMA(0,1,3)	42	0	0	0	28	100	100	100
IMA(0,1,4)	0	16	0	0	11	94	100	100
d=2								
ARI(1,2,0)	22	11	11	0	33	89	100	100
ARI(2,2,0)	14	3	3	17	31	64	100	100
ARI(3,2,0)	11	4	0	7	29	36	100	100
ARI(4,2,0)	22	12	18	24	14	20	100	100
IMA(0,2,1)	0	0	100	100	100	100	100	100
IMA(0,2,2)	10	0	0	0	47	100	100	100
IMA(0,2,3)	42	0	0	0	28	100	100	100
IMA(0,2,4)	0	16	0	0	9	85	100	100

Performance of the ADF test

Table 1 illustrates the percentage performance of the ADF test. For Autoregressive models at all levels, the ADF test's performance at sample sizes 10, 15, 20, 25, 45, and 100 is weak except AR(1) and AR(2) at sample size 100. However, when the sample sizes are 1000 and 10000, the ADF test performs excellently. For the Moving Average models at all levels, the ADF test's performance at sample sizes 10, 15, 20, 25, and 45 was weak except for MA(1) at sample sizes 20, 25, and 45. It performed excellently at sample size 100, 1000, and 10000. Hence, the performance was relatively the same as far as the order of integration is concerned.

Performance of the PP Test

Table 2 illustrates the performance of the PP test. For all cases of the order of integration, AR's performance at lower sample sizes (i.e., 10 and 15) was low. However, sample sizes from 20 onwards performance of the PP test were better. For MA processes at all integration cases, the performance was excellent from sample sizes 100 upward, whereas the other model's performance was better from sample size 45 upwards.

Generally, in both AR and MA models, the PP's performance at small sample sizes is poor. For large sample sizes, the performance was excellent. Hence, with increasing sample sizes, the performance of the PP test improved.

Table 2. Percentage of selecting stationarity using PP

Model	Sample Size							
	10	15	20	25	45	100	1000	10000
d=0								
ARI(1,0,0)	0	0	0	11	56	89	100	100
ARI(2,0,0)	0	11	31	58	78	83	100	100
ARI(3,0,0)	4	32	71	93	100	100	100	100
ARI(4,0,0)	6	14	59	80	86	90	100	100
IMA(0,0,1)	0	0	0	0	44	100	100	100
IMA(0,0,2)	0	17	32	51	100	100	100	100
IMA(0,0,3)	0	9	20	91	100	100	100	100
IMA(0,0,4)	0	16	34	73	100	100	100	100
d=1								
ARI(1,1,0)	0	20	0	11	56	89	100	100
ARI(2,1,0)	0	11	31	58	89	83	100	100
ARI(3,1,0)	4	32	68	93	100	100	100	100
ARI(4,1,0)	0	12	67	94	96	98	100	100
IMA(0,1,1)	0	0	0	0	44	100	100	100
IMA(0,1,2)	0	17	32	51	100	100	100	100
IMA(0,1,3)	0	9	20	91	100	100	100	100
IMA(0,1,4)	0	16	34	73	100	100	100	100
d=2								
ARI(1,2,0)	0	0	0	11	56	89	100	100
ARI(2,2,0)	0	11	31	58	89	83	100	100
ARI(3,2,0)	4	32	68	93	100	100	100	100
ARI(4,2,0)	0	16	67	92	96	98	100	100
IMA(0,2,1)	0	0	0	0	56	100	100	100
IMA(0,2,2)	0	17	33	51	100	100	100	100
IMA(0,2,3)	0	9	20	91	100	100	100	100
IMA(0,2,4)	0	16	34	73	100	100	100	100

Table 3. Percentage of selecting stationarity using KPSS

Model	Sample Size							
	10	15	20	25	45	100	1000	
d=0								
ARI(1,0,0)	100	78	100	100	89	100	100	100
ARI(2,0,0)	100	97	92	83	89	86	100	100
ARI(3,0,0)	100	89	71	64	64	50	100	100
ARI(4,0,0)	100	100	100	100	100	100	100	100
IMA(0,0,1)	100	100	100	100	100	100	100	100
IMA(0,0,2)	100	100	100	100	100	100	100	100
IMA(0,0,3)	100	100	100	100	100	100	100	100
IMA(0,0,4)	100	100	100	100	100	100	100	100
d=1								
ARI(1,1,0)	78	78	100	100	89	100	100	100
ARI(2,1,0)	100	97	92	83	89	86	100	100
ARI(3,1,0)	100	89	71	64	64	50	100	100
ARI(4,1,0)	100	100	100	100	100	100	100	100
IMA(0,1,1)	100	100	100	100	100	100	100	100
IMA(0,1,2)	100	100	100	100	100	100	100	100
IMA(0,1,3)	100	100	100	100	100	100	100	100
IMA(0,1,4)	100	100	100	100	100	100	100	100
d=2								
ARI(1,2,0)	100	78	100	100	100	89	100	100
ARI(2,2,0)	100	97	92	83	78	75	100	100
ARI(3,2,0)	100	89	64	64	64	50	100	100
ARI(4,2,0)	100	100	100	100	100	100	100	100
IMA(0,2,1)	100	100	100	100	100	100	100	100
IMA(0,2,2)	100	100	100	100	100	100	100	100
IMA(0,2,3)	100	100	100	100	100	100	100	100
IMA(0,2,4)	100	100	100	100	100	100	100	100

Performance of the KPSS Test

Table 3 illustrates the performance of the KPSS test. The KPSS test's performance on the MA processes was excellent in all the sample sizes at all integration levels. However, for the AR (2) and AR(3) at all levels of integration, there was a decreasing trend in performance as sample sizes increased from 20 to 100. The performance, in this case, was excellent for 10, 1000, and 10000. Hence, the order of integration $d=0$, $d=1$, and $d=2$ does not affect the KPSS test's performance with regards to the MA processes.

Effect of increase in the order of the model on the three conventional tests of stationarity

Table 4 illustrates the average percentage performance as the order of the model's increases. As the order of both Autoregressive and Moving Average at level increases from order 1 to 4, the ADF test's performance was not ordered and was low. Still, for large sample sizes, the performance decreased as the model's order increases, which shows a negative relation. As the order of the AR models increases from order 1 to 4, the performance of the PP was not ordered for both small and large sample sizes. However, the PP's performance increases as the Moving Average increases from order 1 to 4 for both small and large sample sizes. The KPSS test performed excellently in both Autoregressive and Moving Average models for small and large sample sizes. Again, the order of integration does not affect the three most commonly used unit root tests as far as increasing the model's order is concerned.

Table 4. Increasing the Order of the Model

Test	Model	d=0		d=1		d=2	
		Small	Large	Small	Large	Small	Large
ADF	AR(1)	11	80.5	11	80.5	11	80.5
	AR(2)	7.75	73.75	8.5	73.75	9.25	73.75
	AR(3)	4.75	66.25	7.75	68.75	5.5	66.25
	AR(4)	19.5	58.5	19.5	58.5	19	58.5
	MA(1)	50	100	50	100	50	100
	MA(2)	4.5	86.75	14.25	86.75	2.5	86.75
	MA(3)	10.5	82	17.5	82	10.5	82
	MA(4)	3.75	75.5	6.75	76.25	4	73.5
PP	AR(1)	2.75	86.25	7.75	86.25	2.75	86.25
	AR(2)	25	90.25	25	93	25	93
	AR(3)	50	100	49.25	100	49.25	100
	AR(4)	39.75	94	43.25	98.5	43.75	98.5
	MA(1)	0	86	0	86	0	100
	MA(2)	25	100	25	100	25.25	100
	MA(3)	30	100	30	100	30	100
	MA(4)	30.75	100	30.75	100	30.75	100
KPSS	AR(1)	94.5	97.25	89	97.25	94.5	97.25
	AR(2)	93	93.75	93	89	93	88.25
	AR(3)	81	78.5	81	76.75	79.25	78.5
	AR(4)	100	100	100	100	100	100
	MA(1)	100	100	100	100	100	100
	MA(2)	100	100	100	100	100	100
	MA(3)	100	100	100	100	100	100
	MA(4)	100	100	100	100	100	100

Effect of sample size on the three unit root tests

Table 5. illustrates the average percentage values of the unit root tests as the order of the model's increases. With the ADF test, as the sample size increased from 10 to 20 concerning AR and MA models at level, the performance decreased and started to increase when the sample sizes were 25 to 10000. Hence, the performance at sample size 1000 and 10000 was excellent. Unlike the ADF and the PP tests have an increasing effect on the performance as sample size increase from 10 to 10000 in both AR and MA models at the level. In the KPSS, as sample size increased from 10 to 10000, the performance decreased in the Autoregressive models at a level from sample size 10 to 100 and became excellent when the sample sizes were 1000 and 10000. The performance of the KPSS in the Moving Average models at level was perfect at all the sample sizes. Moreover, when the series was differenced once and twice that is $d = 1, 2$ there was no change in the three unit root tests' performance.

The overall performance of the three conventional unit root tests

Table 6 illustrates the overall performance of the three

most commonly used unit root tests. The performance in percentages of the unit root test were grouped in classes, PG with weights assigned to each of the classes. The classes were 0-20 with an assigned weight of 0, 21-41 with an assigned weight of 1, 42-62 with an assigned weight of 2, 63-83 with a given weight of 3, and 84-100 with a given weight of 4. The frequencies, r_i in each of the performance groups were multiplied by their respective weights, ω_i and the results were summed together to obtain the Score for each of the unit root tests.

The Score is calculated as:

$$Score = \sum \omega_i r_i$$

$$Score_{ADF} = 0 * 87 + 1 * 24 + 2 * 6 + 3 * 3 + 4 * 72 = 333$$

$$Score_{PP} = 0 * 6 + 1 * 12 + 2 * 13 + 3 * 13 + 4 * 94 = 419$$

$$Score_{KPSS} = 0 * 0 + 1 * 1 + 2 * 5 + 3 * 32 + 4 * 154 = 723$$

Based on the Scores in Table 6, the KPSS test was ranked as the overall best performing test compared to ADF and PP tests. This was followed by the PP test and then lastly the ADF test.

Table 5. Increasing Sample size

Test	Model	10	15	20	25	45	100	1000	10000
d=0									
ADF	AR	18.5	7.5	7.25	9.75	26.75	52.25	100	100
	MA	13.25	3.75	25	26.75	46	98.25	100	100
PP	AR	2.5	14.25	40.25	60.5	80	90.5	100	100
	MA	0	10.5	21.5	53.75	86	100	100	100
KPSS	AR	100	91	90.25	86.75	85.5	84	100	100
	MA	100	100	100	100	100	100	100	100
d=1									
ADF	AR	17.75	7.5	7.25	8	26.75	54.75	100	100
	MA	13	4	25	25	46.5	98.5	100	100
PP	AR	1	18.75	41.5	64	85.25	92.5	100	100
	MA	0	10.5	21.5	53.75	86	100	100	100
KPSS	AR	94.5	91	90.75	86.75	82.75	82	98.25	100
	MA	100	100	100	100	100	100	100	100
d=2									
ADF	AR	17.75	7.5	8	12	26.75	52.25	100	100
	MA	13	4	25	25	46	96.25	100	100
PP	AR	1	14.75	41.5	63.5	85.75	92.5	100	100
	MA	0	10.5	21.75	53.75	89	100	100	100
KPSS	AR	94.5	91	89	86.75	85.5	78.5	100	100
	MA	100	100	100	100	100	100	100	100

Table 6. Overall Performance of the Unit Root Tests

Weight(ω_i)	0	1	2	3	4	SCORE	RANK
PG(%) [r_i]	0-20	21-41	42-62	63-83	84-100		
ADF	87	24	6	3	72	333	THIRD
PP	60	12	13	13	94	419	SECOND
KPSS	0	1	5	32	154	723	FIRST

4. Conclusions

In summary, the issue of unit root tests has recently grown in importance. Most studies have evaluated various unit root tests. The commonly used unit root tests are the ADF, PP, and KPSS tests. Nonetheless, primarily analysis done by several authors has shown disagreement with these unit root tests. This study compared the performances of the unit root tests through a simulation study.

To compare the performances of these three unit root tests, a simulation study was performed, out of which 24 different univariate time series models and 8 different sample sizes were used. Regarding the unit root tests' performances under different conditions, our results showed that the KPSS was exceptional compared to the ADF and PP considering the overall performance.

The ADF and PP performances were poor for small sample sizes but were excellent for very large sample size

in all the differencing orders. Also, the KPSS test's performance was excellent for all the sample sizes in all the differencing orders.

Increasing the order of both AR and MA models decreases the ADF test's performance for a large sample size but unordered for a small sample size. The performance of the PP test in AR models is unordered for both small and large sample sizes, but MA models increase as the order increases for both small and large sample sizes. The performance of the KPSS is also unordered as model order increases in the AR models and excellent in MA models as the order increases.

As the sample size increases, all three tests' performance increases in both AR and MA models.

In comparison to the other two unit root tests, the KPSS had an astounding performance.

REFERENCES

- [1] Tebbs, J. M. Lecture notes in forecasting and time series, Department of Statistics University of South Carolina, 2013.
- [2] Adhikari, R. and Agrawal, R. K. An introductory study on time series modeling and forecasting, *arXiv preprint arXiv:1302.6613*, 2013.
- [3] Test, D.-F. Deterministic or stochastic trend, *Methodology* 6(2): 83-92, 2010.
- [4] Fedorov'a, D. Selection of unit root test on the basis of length of the time series and value of AR (1) parameter, *Statistika* 96(3): 3, 2016.
- [5] Twumasi-Ankrah, S. Lecture notes in time series and forecasting, Unpublished Lecture Note, University of Mines and Technology, 101 pp, 2016.
- [6] Brownlee, J. How to check if time series data is stationary with python, *Machine Learning Mastery*, 2016.
- [7] Masliah, M. R. Stationarity/nonstationarity identification, Internet Publication, 2002.
- [8] Cochrane, J. H. A critique of the application of unit root tests, *Journal of Economic Dynamics and Control* 15(2): 275-284, 1991.
- [9] Mushtaq, R. Augmented dickey fuller test, 2011.
- [10] Davidson, R., & MacKinnon, J. G. *Econometric theory and methods* (Vol. 5). New York: Oxford University Press, 2004.
- [11] Schwert, G. W. *et al.* Effects of model specification on tests for unit roots in macroeconomic data, *Journal of monetary economics* 20(1): 73-103, 1987.
- [12] Hadri, K. and Rao, Y. Kpss test and model misspecifications, *Applied Economics Letters* 16 (12): 1187-1190, 2009.
- [13] Phillips, P. C. and Jin, S. The kpss test with seasonal dummies, *Economics Letters* 77(2): 239-243, 2002.