ISSN(E):2522-2260
ISSN(P):2522-2252

Indexing/Abstracting

Journal of Quantitative Methods

Detecting Stationarity of GDP: A Test of Unit Root Tests

Author(s)
Atiq-ur-Rehman[1]

Affiliations
[1]Department of Econometrics and Statistics, Pakistan Institute of Development Economics, Islamabad, Pakistan.

Manuscript Information

Submission Date: May 01, 2018
Acceptance Date: February 27, 2019

Citation in APA Style

Rehman, A. (2019). Detecting stationarity of GDP: A test of unit root tests, *Journal of Quantitative Methods*, *3*(1), 8-37.

This manuscript contains references to 23 other manuscripts.

The online version of this manuscript can be found at
https://journals.umt.edu.pk/sbe/jqm/volume3issue1.aspx#

DOI: https://doi.org/10.29145/2019/jqm/030102

Published by

Department of Quantitative Methods
School of Business and Economics
University of Management and Technology, Lahore, Pakistan

Additional Information

Subscriptions and email alerts: editorasst.jqm@umt.edu.pk
For further information, please visit
http://journals.umt.edu.pk/sbe/jqm/Home.aspx

Detecting Stationarity of GDP: A Test of Unit Root Tests

Atiq-ur-Rehman[1]

https://doi.org/10.29145/2019/jqm/030102

Abstract

Despite extensive research of research on unit roots, consensus on several important issues and implications has not emerged to date (Libanio, 2005). There are many series which were being investigated for existence of unit root and for these series, there is conflict between the researchers regarding the existence of unit root. For a given data series it is generally not possible to decide which of unit root tests would be the best suited. The Monte Carlo experiments prove that the performance of unit root tests depends on the type of data generating process (DGP), but for the real data we do not know the true DGP. Hence, we cannot decide which of the tests would perform best for a series. The bootstrap approach of Rudebusch (1993) offers an alternative to measure the performance of unit root test for any real time series with unknown DGP. Rudebusch (1993)'s approach is extended to measure and compare the performance of unit root tests for annual real GDP series of various countries. Our results show that unit root tests have very low ability to discriminate between best fitting trend stationary and difference stationary models for GDP series of most of the countries and that Phillips Perron test is superior to its rivals including Dickey-Fuller, DF-GLS and Ng-Perron tests. The results also support existence of unit root in real GDP series.

Keywords: unit root tests, stationarity, GDP

JEL Classifications: C01, C15, C22

[1]Department of Econometrics and Statistics, Pakistan Institute of Development Economics, Islamabad, Pakistan.
E-mail: ateeqmzd@gmail.com

Author's Acknowledgment: I (Atiq-ur-Rehman), hereby, declare that the given paper is extracted from my PhD thesis and/or is an extension of my previous research work.

1. Introduction

In a seminal paper, Nelson and Plosser (1982) showed that several common economic time series had stochastic, rather than deterministic, trends. These two statistical specifications are radically different both in terms of statistical and in terms of economic implications. Unit root tests are the principal means of discriminating between the two models, and a huge literature has developed since then. For a recent survey, see Patterson (2010).

An extremely large number of unit root tests have been proposed, and very little guidance is available regarding relative performance of these tests. Huge survey of these tests and their comparisons exist e.g. Maddala and Kim (1998) and Perron (2006). However, these do not resolve the problem, since different tests have different areas of strengths and weaknesses. For example, a test that is designed to test unit root in presence of structural breaks would be better when there are structural breaks and will lose its desirable properties when there is no structural break. On the real data, the performance of these tests cannot be assessed because we don't know what the true data generating process is. Rudebusch (1993)'s bootstrap approach which is summarized in the next sections, offers an alternative to measure the performance of unit root test for any real time series with unknown DGP. Rudebusch's methodology is extended to find the ability of the unit root tests to differentiate between unit root and stationary series and to make a mutual comparison of various unit root tests.

Rudebusch approach is utilized to measure the performance of unit root tests for the GDP series of various countries, and to compare the tests with each other. Results show that for most of data series, unit root tests are unable to discriminate between best fitting trend stationary and difference stationary models. For some series, it is possible to discriminate between two types of models and the Phillips Perron test performs best for the purpose. Our results also support existence of unit root in GDP series.

The rest of the paper is organized such that Section 1 introduces bootstrap approach introduced by Rudebusch (1993) and a discussion that why this approach is suitable to evaluate the performance of unit root tests. Section 2 introduces the modifications made in the Rudebusch approach by the author. Section 3 discusses the computational details of the unit root tests being compared in this paper. Section 4 is about the specification decisions needed for unit root testing and the author's strategy to make these decisions. Section 5 is about the data and sample size used in this paper. Section 6 gives details of parametric spaces estimated from the real data and used further for simulations. Section 7 is about the Monte Carlo experiment and its results. Section 8 contains the discussion on the results. Section 9 discusses the real life implications of the results and finally section 10 concludes the discussion.

2. The Bootstrap Approach of Rudebusch and Comparison of Unit Root Tests

Our aim in this paper is to find answer to two questions: (i) is it possible to discriminate between trend stationary and difference stationary model for GNP series of various countries, the opposite of this could be taken as observational equivalence (ii) if it is possible to discriminate between trend and difference stationary models, which of the unit root tests performs best for the purpose. Since voluminous literature on the unit root already exists, a detailed survey of literature is not much useful. Interested readers are referred to relevant sources including Maddala and Kim (1998) and Patterson (2010). An important limitation of these studies is lack of compatibility with real data. Most of these studies are based on Monte Carlo simulations whereas few comparisons are based on asymptotic properties. Unfortunately, Monte Carlo simulations studies offer us no guidance on which test should be used in real world applications, such as that of finding a unit root in the GNP series. The Monte Carlo studies on performance of unit root tests are based on arbitrary pre-specified data generating process and perform well for same data generating process. But for the real series, we have no prior idea of the data generating process.

Rudebusch (1993) takes a forward step and proposes a procedure which uses the real data to evaluate the performance of unit root tests. Rudebusch (1993) measures the ability of a unit root test to discriminate between the best fitting trend stationary and best fitting difference stationary models estimated from given data series. He estimates best fitting trend stationary model and best fitting difference stationary models for given time series and then takes these two estimated models as DGP for computation of size and power. Therefore, this approach offers systematic procedures for choosing the

DGP instead of arbitrary choice and the procedures provide a model having close matching with the properties of time series under consideration The Rudebusch (1993) approach is outlined as under:

For a given real time series y_t, compute the best fitting trend stationary model by estimating following autoregression:

$$y_t = a + bt + \sum_{i=1}^{k} \phi_i y_{t-k} + \varepsilon_t \quad (1)$$

For the same series, compute the best fitting difference stationary model by estimating following autoregression:

$$\Delta y_t = \alpha + \sum_{i=1}^{k} \gamma_i \Delta y_{t-k} + v_t \quad (2)$$

Use the estimates of a, b, ϕ_i and σ_ε^2 to generate artificial data series analogues to trend stationary (TS) model of the real data series. Compute the unit root test statistics for this series.

Use the estimates of α, γ_i and σ_v^2 to generate artificial data series analogues to difference stationary (DS) model of the real data series. Compute the unit root test statistics for this series.

Repeating the above process for a large number of times one can estimate distribution of the test statistics for two types of models. If the two distributions are distant to each other than the unit root test would be able to discriminate between the two types of models whereas it would fail if major portion of the distributions is overlapping.

3. Extending the Rudebusch Approach

As stated above, Rudebusch (1993) measures the ability of a unit root test to discriminate the best fitting trend stationary and difference stationary models estimated from given data series. Rudebusch (1993) approach is extended in two directions as follows:

i. Rudebusch (1993) procedure measures the performance of single unit root test; we use this approximation of the performance to compare various tests.
ii. Rudebusch (1993) estimates best fitting trend stationary and difference stationary model for single time series and then uses these estimates to evaluate size and power of unit root tests. We formulate

two parametric spaces covering the estimated parameters of the best fitting difference stationary and trend stationary models of a large pool of countries. The performance of unit root tests is evaluated on these parametric spaces. Thus, the results can be generalized to any data series, whose estimated parameters fall into these parametric spaces.

Extensive bootstrap simulation experiments were performed to compute the size and power of various unit root tests for models belonging to the two parametric spaces. Although, the scope of study is limited to the series whose parameters fall into these parametric spaces, our results give a fair measure of the ability of unit root tests to differentiate between trend stationary and difference stationary models.

4. Tests in Competition

In this study, we have utilized four univariate unit root tests: Augmented Dickey Fuller (ADF) test, Phillips Perron (PP) test, Dickey Fuller GLS (DF-GLS) test and Ng-Perron (NP) test. Including their variation with respect to deterministic trend, we have sixteen tests to be compared. The detail on computation of tests statistics and critical values is discussed in detail in is present in next section. The tests are:

- (Augmented) Dickey Fuller Test
  (i) Without drift and trend (DFN), (ii) With drift but no trend (DFC) and (iii) With drift and Trend (DFT)
- Phillips Perron Test
  (i) Without drift and trend (DFN), (ii) With drift but no trend (DFC) and (iii) With drift and Trend (DFT)
- Dickey Fuller GLS tests
  (i) Without Trend (DFGC) and (ii) With Trend (DFGT)
- Ng Perron Test
  (i) MZA without Trend (ZAC), (ii) MZA with Trend (ZAT), (iii) MSB without Trend (SBC), (iv) MSB with Trend (SBT), (v) MPT without Trend (PTC), (vi) MPT with Trend (PTT), (vii) MZT without Trend (ZTC) and (viii) MZT with Trend (ZTT)

a. Augmented Dickey Fuller Test

ADF test is the modified version of test statistics proposed by Dickey and Fuller (1979). ADF test statistics is based on one of following regression equations.

M1	Without drift, trend	$\Delta y_t = \rho y_{t-1} + \sum_{i=1}^{k} \gamma_j \Delta y_{t-k} + e_t$	
M2	With drift, but no trend	$\Delta y_t = \alpha + \rho y_{t-1} + \sum_{i=1}^{k} \gamma_j \Delta y_{t-k} + e_t$	(3)
M3	With drift and trend	$\Delta y_t = \alpha + \beta\, t + \rho y_{t-1} + \sum_{i=1}^{k} \gamma_j \Delta y_{t-k} + e_t$	
	Where	$e_t \sim iid(0, \sigma^2)$	

The test statistics is given by $t_{\hat{\rho}} = \dfrac{\hat{\rho}}{SE(\hat{\rho})}$, where $\hat{\rho}$ is OLS estimate of ρ.

Asymptotic distribution of ADF test statistics is non-standard. Therefore, the critical values are to be computed by simulations or numerical approximations. The critical values of ADF test statistics are provided by McKinnon (1994) computed via Monte Carlo experiments.

b. Phillips–Perron Test

Phillips-Perron test is a unit root test, based on the Dickey-Fuller regression equation. But unlike the Augmented Dickey–Fuller test, which extends the Dickey-Fuller test by including additional lags of variables as regressors in the model, the Phillips-Perron test makes a non-parametric correction to the t-test statistic to capture the effect of autocorrelation.

i. The Phillips Perron Test Statistics

The Phillips Perron test statistics is based on one of the three regression equation describe below:

1	Without drift, trend	$\Delta y_t = \rho y_{t-1} + e_t$	
2	With drift, but no trend	$\Delta y_t = \alpha + \rho y_{t-1} + e_t$	(4)
3	With drift and trend	$\Delta y_t = \alpha + \beta\, t + \rho y_{t-1} + e_t$	
	Where	$e_t \sim iid(0, \sigma^2)$	

These three equations are similar to Dickey Fuller regression equations without any ‘augmentation’. The test statistics is given by:

$$\tilde{t}_{\rho} = t_{\hat{\rho}}\left(\frac{s^2}{\hat{f}(0)}\right) - T\left(\frac{\hat{f}(0) - s^2 \quad SE(\hat{\rho}}{2\left(\hat{f}(0)\sum_{t=2}^{T} e_t^2\right)^{0.5}}\right) \tag{5}$$

where $t_{\hat{\rho}} = \dfrac{\hat{\rho}}{SE(\hat{\rho})}$, $s^2 = T^{-1}\sum_{t=2}^{T} e_t^2$ and e_t are the residuals of the regression. $\hat{f}(0)$ is estimate of spectral density at frequency zero whose estimation procedure is described below. The limiting distributions of Phillips Perron test statistics are similar to corresponding distributions of Dickey Fuller test. Finite sample critical values are also same.

ii. Estimating Spectral Density at Frequency Zero

There are various ways of computing spectral density at frequency zero for a series. Following Ng and Perron (2001), we will use autoregressive estimate of spectral density, wherever needed in the thesis. This can be computed as follows:

Consider the ADF regression equation described in (3). Estimate number of lags included in ADF equation using some consistent criterion e.g. MAIC. Than the estimate of autoregressive spectral density at frequency zero is given by:

$$\hat{f}(0) = \frac{\hat{\sigma}^2}{1 - \hat{\gamma}_1 - \hat{\gamma}_2 - \ldots\hat{\gamma}_k} \tag{6}$$

where $\hat{\sigma}^2$ is estimate of error variance and $\hat{\gamma}_i, i = 1,\ldots k$ are the estimated coefficients from regression equation 3.

c. DF-GLS Test

Elliott, Rothenberg and Stock (1992), use King (1987)'s approach to develop a best point optimal test. They find a test whose power function is tangent to the power envelope and never far below it. Then they find a test which has power function closest to this test. This test is based on GLS detrending whose procedure is as follows:

Let $y_1, y_2, \ldots y_t$ be the data series. The quasi differenced series is obtained as:

$$\nabla y_t = \begin{cases} y_t & \text{if} \quad t = 1 \\ y_t - a y_{t-1} & \text{if} \quad t > 1 \end{cases}. \tag{7}$$

Next considered following OLS regression:

$$\nabla y_t = \nabla x_t \beta + u_t \tag{8}$$

where x_t is the deterministic part; the GLS detrended series y_t^d is defined as:

$$y_t^d = y_t - x_t \hat{\beta} \tag{9}$$

$\hat{\beta}$ is the estimate of β from (8). The deterministic part x_t would be vector of ones, $\{1\} = (1, 1, \ldots 1)'$ if series is assumed not to have linear trend and $\{1, t\} = \begin{pmatrix} 1,1,\ldots..1 \\ 1,2,\ldots\ldots T \end{pmatrix}'$ if series is assumed to have a linear trend. Value of a is chosen as $a = \frac{-13.5}{T}$ if series is assumed to have linear trend $a = \frac{-7}{T}$ if series does not have linear trend. This procedure is also called local to unity GLS detrending. The DF-GLS statistics is then computed from following regression:

$$\Delta y_t^d = \rho y_{t-1}^d + \sum_{i=1}^{k} \gamma_j \Delta y_{t-i}^d + e_t \tag{10}$$

And the test statistics $\hat{t}_{GLS} = \dfrac{\hat{\rho}}{SE(\hat{\rho})}$

Elliott, Rothenberg and Stock (1992), show that the power curve of $\hat{t}_{GLS}$ is tangent to asymptotic power envelop and is never far below it. The finite sample critical values can be found in Elliot at al. (1992).

d. NG-Perron Test

Elliott, Rothenberg and Stock (1992), showed that power function of their test is tangent to power envelop at 50% power. However, inappropriate choice of lag length can still lead to poor size/power properties. While the power gains of the DF using GLS detrended data are impressive, simulations also show that the test exhibits strong size distortions when there is MA root with negative coefficient. Size distortions, however, are less of an issue with the M-tests in theory as shown by Perron and Ng (1996).

In practice, it does require us to have a way to find the appropriate lag length. So, Ng & Perron kept these three things in mind and designed M test for GLS detrended data. They also designed a criterion for choice of appropriate lag length, which they show better than other existing criteria. Therefore, this test accumulates the intellectual wisdom of GLS detrending proposed by Elliot, Rothenberg and Stock (1992), and usage of M-estimators proposed by Stock (1999). M-type test uses the estimate of spectral density of autoregressive process. Ng and Perron (2001) proposed a set of four tests all using M-estimator. Further detail on computation of these tests is as under:

Let $y_1, y_2, \ldots y_t$ be a time series to be tested for unit root. Compute GLS-detrended series $y_1^d, y_2^d, \ldots y_T^d$ as defined in equation 9 Consider the OLS regression equation 10, i.e.

$$\Delta y_t^d = \rho y_{t-1}^d + \sum_{j=1}^{k} \gamma_j y_{t-j}^d + e_{tk}$$

Than spectral density estimate at frequency zero from equation 4 is:

$$\hat{f}(0) = \hat{\sigma}^2 \;\; 1 - \hat{\gamma}_1 - \hat{\gamma}_2 - \ldots \hat{\gamma}_k \;\; ^{-1}$$

Define $\kappa = \sum_{t=2}^{T} (y_{t-1}^d)^2$

The set of tests proposed by Ng and Perron contain tests MZ_α, MZ_t, MSB and MP_T. These tests are defined as follows:

$$MZ_\alpha = \frac{T^{-1}(y_T^d)^2 - \hat{f}(0)}{2\kappa} \tag{11}$$

$$MZ_\alpha = \frac{T^{-1}(y_T^d)^2 - \hat{f}(0)}{2\kappa} \tag{12}$$

$$MZ_t = MSB \times MZ_\alpha \tag{13}$$

$$MP_T = \begin{cases} \dfrac{1}{\hat{f}(0)}\left(a^2\kappa - aT^{-1}\ y_T^{d\ 2}\right) & \text{if} \quad x = \{1\} \\ \dfrac{1}{\hat{f}(0)}\left(a^2\kappa + (1-a)T^{-1}\ y_T^{d\ 2}\right) & \text{if} \quad x = \{1, t\} \end{cases} \tag{14}$$

where a is equal to -7 if $x = \{1\}$ and -13.5 if $x = \{1, t\}$.

i. Asymptotic Behavior and Critical Values of Ng-Perron Test

Ng and Perron claim that the four tests have optimal properties of DF-GLS test and M-estimator proposed by Stock (1999). They argue that asymptotic power curve of these tests is never far below the asymptotic power envelop. The asymptotic critical values of Ng-Perron test are provided by Ng and Perron (2001).

5. Pre-Test Model Specification

Before application of unit root test to a real data series, a researcher has to make number of specification decisions. Two important decisions are the choice of lag length and specification of deterministic regressors. There are various methods for making such decisions and among these methods the methods utilized in this study are summarized below.

a. Criterion for Choice of Lag Length

Appropriate choice of truncation lag is important for the implementation of unit root test proposed by Dickey and Fuller (1979) and Said and Dickey (1984). It is also required to estimate the autoregressive spectral density at frequency zero. Several criteria exist for the choice of truncation lag. Ng and Perron (2001) compare performance of several criteria for the choice of lag length and show that Modified Akaike Information Criterion outperforms other criteria for the appropriate choice of lag length. Following Ng and Perron (2001), throughout this study we will use MAIC for the choice of lag length. This MAIC statistics is given as under:

For the autoregression defined $\Delta\tilde{y}_t = \rho\tilde{y}_{t-1} + \sum_{j-1}^{k}\gamma_j\ \tilde{y}_{t-j} + \varepsilon_t$, the MAIC is computed as:

$$MAIC = \ln(\hat{\sigma}_k^2) + \frac{2\ \tau_T(k) + k}{T - k_{\max}} \tag{15}$$

Here $\hat{\sigma}_k^2$ is the variance of residuals from regression equation 1 when k lags are included in the autoregression and $T_T(k) = \left(\hat{\sigma}_K^2\right)^{-1} \hat{\rho}^2 \sum_{T=K_{max}+1}^{T} \hat{y}_{t-1}^2$. Also $\tilde{y}_t = y_t$ for ADF, PP and PP test whereas $\tilde{y}_t = y_t^d$ for DF-GLS and NP test.

b. Choice of Deterministic Part

Appropriate choice of time trend is very important in unit root testing. Inappropriate choice of deterministic trend leads to substantial power loss (Campbell & Perron, 1991). The existing techniques for specification of deterministic trend do not have reliable size and power properties (see Hacker & Hatemi, 2006 and Rehman & Zaman, 2008). Instead of choosing between different specifications of deterministic trend, we analyze all commonly used specifications of deterministic trend. Therefore the Dickey Fuller test and Phillips Perron test are used with three specification of deterministic part i.e. (i) without drift and trend, (ii) with drift and (iii) with drift and trend. Similarly, we use two specifications of deterministic trend for DF-GLS and Ng-Perron Tests.

6. Data and Sample Size

Our focus in this study is the annual GDP series, which shares several common characteristics. One of the important characteristic is the small sample size. Most developing countries have small amount of macroeconomic data, which can be used for econometric analysis. The WDI database which is perhaps the largest data source for data on developing countries and is published by World Bank, consist of annual time series for various countries. This database has data starting from 1960; therefore, the length of data available today is about 55 observations. However, for many countries, the available length of macroeconomic time series data does not exceed 20 observations.

The problem we have to study, is to decide whether a given GNP series is TS or DS, requires working with small samples. This has important implication because many tests which have good size/power in large/moderate sample sizes, fail to perform well in the small samples.

The data we use are GDP per capita (Constant US$) retrieved from WDI data base. We select the countries for which data is available from 1960 to 2010 and there is no evidence of structural break in this period. The structural break is inspected by applying Chow break point test to the following autoregression: $y_t = \alpha + \beta t + \sum_{i=1}^{3} y_{t-i} + \varepsilon_t$. Here y_t is the log transform of the GDP series. There were 96 countries for which we find full length data series. After discarding the data series with structural breaks we are left with the following 55 countries:

Australia, Austria, Belgium, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Central Africa, Chad, China, Cote d'Ivoire, Denmark, Dominican Rep, Ecuador, Egypt, Finland, France, Greece, Guyana, Honduras, Hong Kong, Iceland, Indonesia, Ireland, Italy, Japan, Kenya, Korea, Luxembourg, Malawi, Malta, Nepal, Netherlands, New Zealand, Nicaragua, Niger, Nigeria, Norway, Pakistan, Portugal, Seychelles, Sierra Leone, Spain, Sri Lanka, Sudan, Sweden, Switzerland, Syria, United Kingdom and Zimbabwe.

7. Estimating Best Fitting Models and Empirical Parametric Spaces

For the GDP of selected 55 countries, best fitting trend stationary and best fitting stationary model were estimated using Rudebusch (1993) approach described in section 3**.** The estimated models have various specifications, however, the simplest and most common trend stationary and difference stationary models were chosen to formulate the parametric spaces. Parametric space Θ_{DS} covers the estimated parameters of DS models and Θ_{TS} covers estimated parameters of TS models.

We report best fitting Difference Stationary models in table 1. The simplest most common DS model was $\Delta y_t = a_0 + \varepsilon_t$, where $a_0 \in (0,.25)$ and $se(\varepsilon_t) \in (0,.027)$. Thus, the two dimensional parametric space for DS models is:

$$\Theta_{DS} = \{(a_0, \sigma_\varepsilon^2) : a_0 \in (0, 0.025), \sigma_\varepsilon \in (0, .027)\} \tag{16}$$

This parametric space covers best fitting models for 22 out of 55 countries. Best fitting Trend Stationary models are reported in table

2. The simplest most common TS model was $y_t = a_1 + b_1 y_{t-1} + \zeta_t$, where $a_1 \in (0,.45), b_1 \in (0.85,1)$ and $se(\zeta_t) \in (0,0.3)$. Thus, the parametric space is:

$$\Theta_{TS} = \{(a_1, b_1, \sigma_\zeta^2) : a_0 \in (0,.45), b_1 \in (.85,1), \sigma_\zeta \in (0,.027)\} \quad (17)$$

This parametric space covers best fitting models for 21 out of 55 countries. The intersection covers 9 countries.

8. Monte Carlo Design and Results

a. Monte Carlo Design

Parametric space for DS models i.e. Q_{DS} was divided into multidimensional grid. Each point of this grid was used as parameter of data generating process. Size of unit root tests was computed for the series thus generated at each point of this grid. The parametric space for TS models Q_{TS} was also divided into another multidimensional grid and power of unit root tests was computed at each point of this grid.

b. Size of Tests

Size of various unit root tests is reported in table 3. We see that for all tests, the empirical size does not exceed the nominal size. Therefore, the probability of type I error is bounded above by the nominal size. No distortion of size was observed. Also it was observed that the size of tests is independent from the variance of error term σ^2.

c. Power of Tests

The powers of various unit root tests are reported in table 4 that shows many unexpected results. Most surprising was the failure of tests based on GLS detrending including the Ng-Perron and the DF-GLS test. The DF-GLS is shown to have power closest to asymptotic power envelope (Elliot, Rothenberg & Stock, 1992). Ng-Perron test is a test accumulating intellectual heritage of the DF-GLS test and M-estimator by King (1987). However, the optimality of these tests is based on asymptotic properties.

The simulations show that optimality does not hold for small samples. For instance, the minimum sample size used for simulations by Ng and Perron (2001) is 100, whereas our sample size is 50. Anyway, these simulations show clear superiority of Dickey Fuller and

Phillips Perron tests over the DF-GLS and the Ng-Perron tests in small samples. In fact an overview of Table 4 reveals that the power of detrending based tests i.e. the DF-GLS and Ng-Perron test rarely exceeds their size, so that these tests have no ability to discriminate between the trend and difference stationary processes for data under consideration. Furthermore, an overview of power of tests tabulated in table 4 reveals that ranking of tests according to average power for TS models is as follows: PPC, DFC, PPT and DFT.

d. The Response Surface for Power of Tests

The PPC test and DFC tests have maximum average power for the TS models, thus they have best overall performance in the context under consideration. The response surface function was estimated to decide better test among these two. The response surfaces for DFC and PPC tests are given in figure 1(a & b).

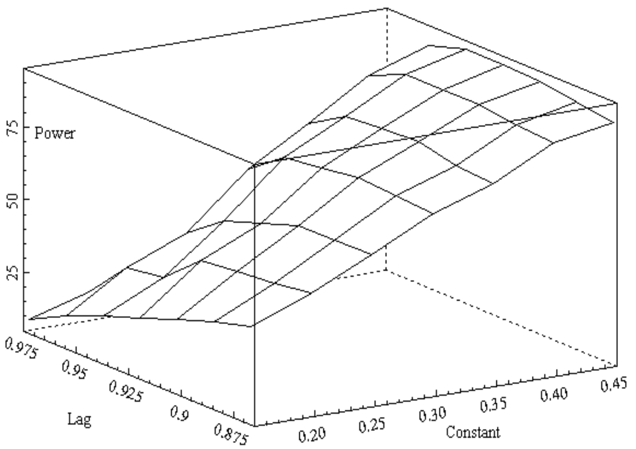

Figure 1a: Response Surface for DFC

The figure gives response surface for DFC test. The power is positively related to distance between unity and lag coefficient, and to the value of constant.

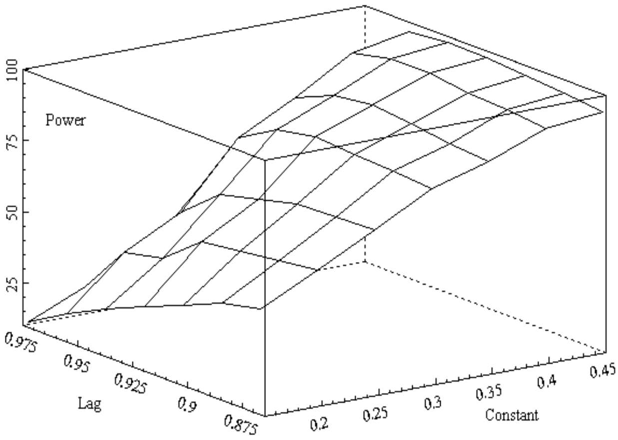

Figure 1b: Response Surface for PPC

The figure gives response surface for PPC test. Just like DFC test, the power is positively related to distance between unity and lag coefficient and to the value of constant.

The response surfaces for the powers of two tests show similar behavior. The power of the tests is positively related to the difference of lag coefficient b_1 from unity i.e. its power increases if the value of b_1 goes to zero (distance from unity increases).

The power is positively associated with the constant a_1 i.e. increases with the increase in value of a_1. Moreover, it can be observed from table 4 that power of PPC test is higher than that of DFC test for entire parametric space.

We compute the approximate response surface functions for the powers of two tests by regressing the power of tests on various functions of $a_1 \quad and \quad b_1$. These response surface functions are:

$$\ln(P_{dfc} \left| a_1, b_1 \right.) = 374.394 - 17.392 a_1 - 211.413 b_1 - 8.03 a_1^2 + 28.765 a_1 b_1 - 163.386 b_1^{-1}$$

And

$$\ln(P_{ppc} \left| a_1, b_1 \right.) = 355.046 - 22.041 a_1 - 201.320 b_1 - 10.032 a_1^2 + 34.762 a_1 b_1 - 154.032 b_1^{-1}$$

where P_{dfc} and P_{ppc} are the powers of DFC and PPC tests respectively. The two models are fairly similar to each other and both provide equal degree of fitness (R-square ≅ 92% for the two models).

The numerical evaluation of the two functions reveals that value of difference $\ln(P_{ppc}|\theta) - \ln(P_{dfc}|\theta)$ is never smaller than zero for all $\theta \in \Theta_{TS}$.

Figure 2 plots the difference between power of PPC test and DFC test i.e. $Diff = P_{ppc} - P_{dfc}$ estimated by using response surface function. Figure 2 confirms that power of PPC test is superior to that of DFC test, since the difference is always positive.

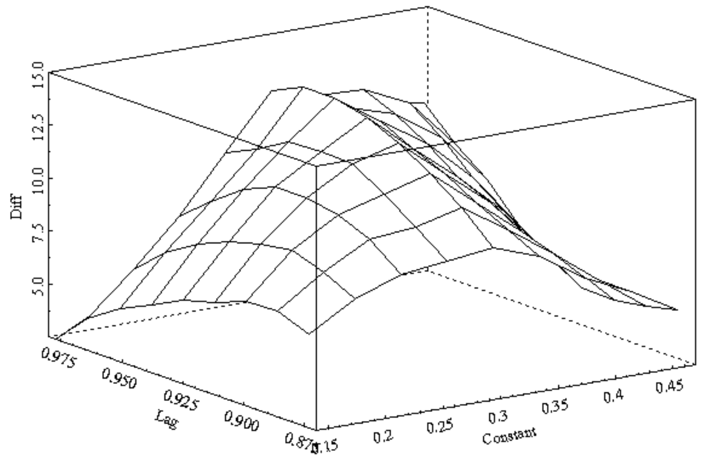

Figure 2: Difference between Powers of PPC and DFC

The figure plots the difference $Diff = P_{ppc} - P_{dfc}$. The difference is positive for all points in parametric space Θ_{TS} which shows that PPC test is superior to DFC test with regard to its power.

The estimated function was then used to predict power of tests for actual models for the real data.

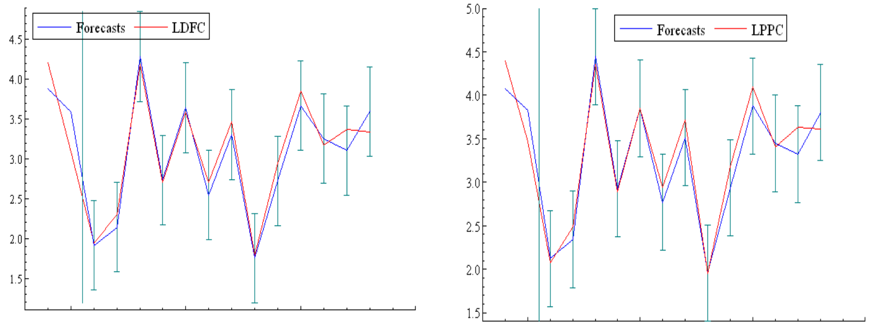

Figure 3: The Predictions by Estimated Response Surface Functions of PPC and DFC

The estimation of Response Surface function was carried out using power of tests at regular grid and this function was then used to predict power of tests on some other points in the parametric space which corresponds to estimated models for real time series. The predictive performance of two tests seems reasonable.

Figure 3 gives the power of PPC and DFC tests for the estimated best fitting models for various countries. It is clear that the PPC test has better performance than DFC for all models. The powers of all other tests are much smaller than the powers of these two tests.

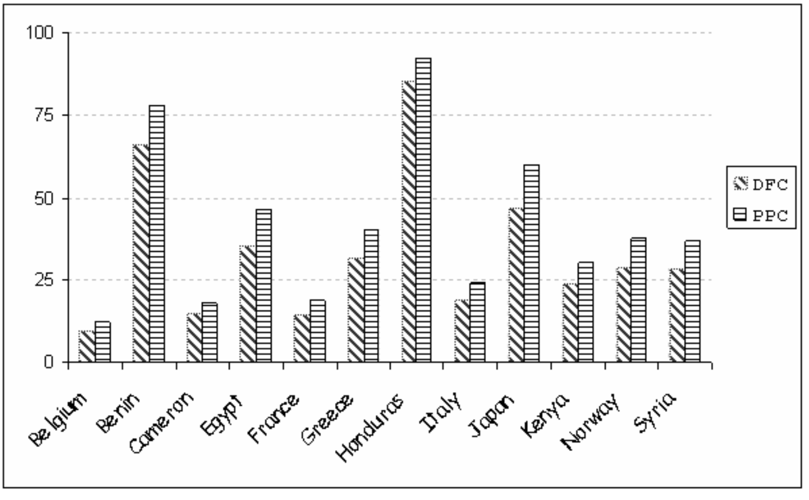

Figure 4: Power of PPC and DFC for Best Fitting TS Models

Powers of DFC and PPC for TS models of various countries are plotted. The superiority of PPC to DFC is clearly visible

It can be seen that for all of the countries, the performance of PPC test is superior to that of DFC test. This leads to the conclusion that the PPC test is superior to other tests with regard to its power for testing stationarity of real GDP series. This superiority occurs without any distortion in the size of test; therefore, the PPC test is superior to all other tests.

9. Discussion of Results

9.1. Observational Equivalences and Reliability of Unit Root Tests

From table 4 (in appendix), we see that most of tests don't have ability to discriminate between trend and difference stationary models that have closer resemblance with the best fitting models for GNP series. The inability to discriminate between trend and difference stationary could be taken as observational equivalences, so if we measure observational equivalence with DF-GLS or Bg-Peron test, the observational equivalence is closer to perfect, and the probability to discriminate between two competing models is closer to zero.

The results show that there are two tests which perform relatively better. The expected power of the best performing test i.e. the PPC tests for various countries based on response surface function is summarized in table 5 (appendix). The simulation results are reported in Figure 3 reveal that actual power of unit root tests does not deviate much from this approximation. Power of PPC test shows different characteristic for different models.

The TS models for various countries can be divided into three groups with respect to the power attained. For first group of countries, say Group I, PPC test has very low probability of rejecting unit root. This group contains the countries for which value of lag coefficient b_1 is close to unity and/or value of drift coefficient a_1 is close to zero. These countries include Malta, Nicaragua, Austria, Belgium, Guyana, Italy and Cameroon. For these countries the PPC test has less than 25% power. Since all other tests have power smaller than PPC, all unit root tests are unable to discriminate between best fitting models of two types for these countries. For these countries, it could be said that the observational equivalence is about 75% or more.

Group II contains the models for which the power of PPC tests is between 25% - 75%. This means the probability of type II error would be also between 25% - 75%. So, the output of PPC test is uncertain for this group of countries. This group includes Norway, Sierra Leone, Kenya, Greece, Zimbabwe, Japan, Syria and Ecuador. There are moderate chances of observational equivalence, if it is to be measured by PP test.

It can also be noted that the PPC test has reasonable power for few countries belonging to Group III. These are the countries with lag coefficient b_1 distant from unity and/or the value of drift coefficient a_1 distant from zero. This third group of countries includes Burundi, Chad, Malawi, Benin and Nigeria and power of PPC test for these countries is more than 75%. This implies that PPC test has reasonable ability to discriminate between trend and difference stationary models for these countries.

For the first two groups, the conclusions of these simulation experiments are similar to the conclusion of Rudebusch (1993), i.e. 'we don't know'. The empirical distribution of trend and difference stationary are so closer to each other that even the best performing test doesn't have enough power to discriminate between two types of models. For counties belonging to Group I & II, all tests including PPC and DFC have the probability of type II error greater than 25%. For the few countries belonging to Group III, only DFC and PPC have reasonable probability to discriminate between trend and difference stationary models. Therefore, the output of unit root tests is not much helpful to discriminate between trend and difference stationary models.

9.2. Comparison of Unit Root Tests

Assume that for GDP of any country, the estimated best fitting trend stationary and difference stationary model are only two possible models. If the true data generating process was difference stationary, the tests should not reject unit root. Table 3 (appendix) gives simulated probabilities of rejection of unit root for the DS models. It can be seen that the probability of rejection of unit root (Type I error) does not exceed 5% nominal size if the estimated parameters lie within the parametric space Θ_{DS}. Therefore, all unit root tests have capability of transmitting right message about stationarity of the series when true model is DS with parameters belonging to the parametric space.

Now if the true data generating process was trend stationary, than the unit root should be rejected. However, table 4 (appendix) reveals that the GLS detrending based tests including DF-GLS and Ng-

Perron test are unable to reject unit root for the trend stationary models with parameters belonging to Q_{TS}. Detrending based tests have the tendency of not rejecting unit root, regardless of the type of data generating process. This means, these tests are unable to determine the type of stationarity for the data under consideration. Similarly DFN and PPN tests are also unable to reject unit root when true DGP is Trend Stationary. The PPT test and DFT tests also have low probability to reject unit root for trend stationary DGP.

However, PPC and DFC tests have maximum probabilities of rejecting unit root if the data was actually generated by TS model. Section 3 reveals that overall best performer test is PPC test.

The power of PPC test depends on the two parameters if the estimated model is generated from parametric space Q_{TS}. Power depends on distance from the unity $1\text{-}b_1$ and on the lag coefficient a_1. Larger values of $1\text{-}b_1$ and a_1 lead to increased power (see Figure 2) and positively related to the value of drift coefficient.

9.3. Stationarity of GDP Series

The analysis presented in 9.2 shows that the tests would be inconclusive for most of the countries. However, for Group III of countries containing Burundi, Chad, Malawi, Benin and Nigeria, we can determine the stationarity of data series with reasonable level of certainty using PPC test. Also for countries belonging to Group II, PPC test has power between 25%-75%. When the unit root tests were applied to real data, all tests failed to reject unit root, for all of the countries included in Group III. This implies the real data series have more resemblance with the DS model.

10. Applications

The discussion presented above reveals that in the time series with smaller sample sizes, the Ng-Perron test and the DF-GLS test have little probability to reject unit root and thus unable to discriminate between the trend and difference stationary model. At the same time Phillips Perron and ADF test do better job to discriminate trend and difference stationary model. Therefore, we predict that Ng-Perron and DF-GLS test will accept null hypothesis of unit root for time series of with small sample sizes. There are number of evidences to support this claim. We provide here some evidences from published results.

Shahbaz, Ahmad and Chaudhary (2008) analyze real GDP per capita, financial development, foreign direct investment, GDP, and annual inflation for Pakistan. Hye, Shahbaz and Butt (2008) analyze output, agricultural terms of trade and technology in agriculture, Hye and Riaz (2008) analyze energy consumption and economic growth for Pakistan using Ng-Perron test. Unit root null was not rejected for all of the series analyzed in three studies.

Sari and Soytas (2007) apply various unit root test to the following Turkish economic timer series: total employment in manufacturing, total electricity consumption in industry, value added-GNP manufacturing and total fixed investment in manufacturing. They apply DF, DFGLS, PP and Ng-Perron test to these series with two specifications of deterministic part i.e. including linear trend and without including linear trend. Their results are totally consistent with the results we computed and summarized. Phillips Perron test reject unit root for some of these series at 1% significance level but Ng Perron test and DF-GLS fail to reject unit root for the same series at 10% level of significance. For the remaining series, neither PP test nor remaining tests reject unit root.

11. Conclusions

A major problem in the comparison of various unit root tests is the absence of information about the data generating process of time series in hand. The properties of unit root tests crucially depend on the DGP, and for the real data, we have no information about the true DGP. The estimation of DGP via general to simple methodology is also not feasible since the performance of estimators depend on existence or otherwise of unit root.

Rudebusch's (1993) approach offers an alternative to measure the performance of unit root test for any given series with unknown DGP. Rudebusch (1993) first estimates best fitting trend stationary and difference stationary models. The two models provide unbiased and consistent estimates of the parameters in general to simple specification procedure since they involve the stationary regressors.

Rudebusch (1993) approach is extended in various dimensions to use it to compare the unit root tests. This procedure gives fairly clear comparison of various unit root tests in terms of their size and power properties.

The findings of this study are summarized as under:

a. **Size of Unit Root Tests:** If we look at the size of various unit root tests, it appears that actual size of all tests is smaller than the nominal size. This means that there is upper bound on probability of Type I error. No size distortion was observed for any of the tests.
b. **Power of Detrending based Unit Root Tests:** The simulated power of unit root tests gives some unexpected results. The most important observation is the failure of tests based on GLS detrending i.e. the DF-GLS and the Ng-Perron tests. DF-GLS test is assumed to have power closest to asymptotic power envelope and the Ng-Perron tests accumulates over the DF-GLS. But it seems that the optimality properties of these tests are based on asymptotic results and our study shows that these properties are not valid for small samples.
c. **Power of ADF and PP Tests:** An overview of power of various unit root tests (Table 4) reveals that the clear winners in competition of unit root tests are PPC tests and DFC tests. The response surface analysis (Section 3) reveals that PPC test is superior to DFC test for all points in the parametric spaces Θ_{TS}.
d. **Reliability of Unit Root Tests:** The simulation results show that most of the tests have tendency to accept unit root even if series is generated by TS model. Only PPC and DFC test have reasonable power for TS models of few countries. Therefore the tests have little ability to discriminate between TS and DS models.
e. **Stationarity of GDP:** The conclusion (d) above shows that the tests would be inconclusive for most of the countries and for few countries we can determine the stationarity of data series with reasonable level of confidence using PPC test. We find that unit root cannot be rejected for any of these countries. Thus it can be concluded that the real GDP series are better described by a DS model. Unit root was also not rejected for the group of countries for which PPC test has power between 25% and 75%.
f. **Limitations of Study:** The limitations of this analysis are presented as under: This analysis is valid if the estimated parameters of best fitting DS and TS models of a series fall within the parametric spaces Θ_{DS} and Θ_{TS}. Also the length of time series was 53 throughout this analysis and results may not hold for longer time series.

References

Campbell, J. Y., & Perron, P. (1991). Pitfalls and opportunities: What macroeconomists should know about unit roots. *NBER Macroeconomics Annual*, *6*, 141-201. https://doi.org/10.2307/3585053.

Dickey, D. A. (1976). *Estimation and hypothesis testing in nonstationary time series*(Doctoral dissertation, Iowa State University), Retrieved from https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=7266&context=rtd.

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. https://doi.org/10.1080/01621459.1979.10482531.

Elliott, G., Rothenberg, T. J., & Stock, J. H. (1992). Efficient tests for an autoregressive unit root. *Econometrica*, *64*(4), 813-36

Hacker. R. &. A. Hatemi-J (2006). *The Properties of Procedures Dealing with Uncertainty about Intercept and Deterministic Trend in Unit Root Testing* (unpublished manuscript).

Hye, Q. M. A., & Riaz, S. (2008). Causality between Energy Consumption and Economic Growth in case of Pakistan, *Lahore Journal of Economics*, *13*(2), 45-58.

Hye, Q. M. A., Shahbaz, M., & Butt, M. S. (2008). Growth in agricultural terms of trade and aggregate supply response model: A case study for Pakistan. *Journal of Agricultural Research*, *46*(4), 411-423.

King, M. L. (1987). Towards a theory of point optimal testing. *Econometric Reviews*, *6*(2), 169-218. https://doi.org/10.1080/07474938708800129.

Libanio, G. A. (2005). Unit roots in macroeconomic time series: Theory, implications, and evidence. *Nova Economia*, *15*(3), 145-176. http://dx.doi.org/10.1590/S0103-63512005000300006.

MacKinnon, J. G. (1994). Approximate asymptotic distribution functions for unit-root and cointegration tests. *Journal of Business & Economic Statistics*, *12*(2), 167-176.

Maddala, G. S., & Kim, I. M. (1998). *Unit roots, cointegration, and structural change* (No. 4). Cambridge, United Kingdom: Cambridge University Press.

Nelson, C. R., & Plosser, C. R. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics*, *10*(2), 139-162. https://doi.org/10.1016/0304-3932(82)90012-5.

Ng, S., & Perron, P. (2001). LAG length selection and the construction of unit root tests with good size and power. *Econometrica*, *69*(6), 1519-1554. https://doi.org/10.1111/1468-0262.00256.

Patterson, K. (2010). *A primer for unit root testing*. UK: Palgrave.

Perron, P. (2006). Dealing with structural breaks. *Palgrave handbook of econometrics*, *1*(2), 278-352.

Perron, P., & Ng, S. (1996). Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. *The Review of Economic Studies*, *63*(3), 435-463.

Perron, P., & Phillips, P. C. (1987). Does GNP have a unit root?: A re-evaluation. *Economics Letters*, *23*(2), 139-145. https://doi.org/10.1016/0165-1765(87)90027-9.

Rehman, A. & A. Zaman, (2008). *Model Specification, Observational Equivalence and Performance of Unit Root Tests*, presented at Far Eastern and South Asian Meeting of Econometric Society, Singapore Management University, Singapore.

Rudebusch, G. D. (1993). The uncertain unit root in real GNP. *The American Economic Review*, *83*(1), 264-272.

Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, *71*(3), 599-607.

Sari, R., & Soytas, U. (2007). The growth of income and energy consumption in six developing countries. *Energy Policy*, *35*(2), 889-898. https://doi.org/10.1016/j.enpol.2006.01.021.

Shahbaz, M., Ahmad, K., & Chaudhary, A. R. (2008). Economic growth and its determinants in Pakistan. *The Pakistan Development Review*, *47*(4), 471-486.

Stock, J. H. (1999). A class of tests for integration and cointegration. *Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive WJ Granger*, 137-167.

ANNEXURE

Table 1: Best Fitting Difference Stationary Models for Various Countries

Country	Const	Lag 1	Sigma	Country	Const	Lag 1	Sigma
Australia	0.009		0.009	Japan	0.005	0.660	0.011
Austria	0.012		0.008	Kenya	0.006		0.018
Belgium	0.011		0.008	Korea	0.025		0.015
Benin	0.002		0.013	Luxembourg	0.013		0.014
Botswana	0.008	0.734	0.013	Malawi	0.004		0.024
Burkina Faso	0.006	-0.332	0.013	Malta	0.007	0.673	0.013
Burundi	0.001		0.024	Nepal	-0.313	-0.313	0.012
Cameroon	0.004		0.026	Netherlands	0.006	0.363	0.007
Central Africa	-0.004	-0.076	0.018	New Zealand	0.006		0.013
Chad	0.001	0.098	0.037	Nicaragua	-0.002	0.301	0.029
China	0.020	0.340	0.020	Niger	-0.007	0.054	0.027
Cote d'Ivoire	-0.001	0.297	0.022	Nigeria	0.003	0.325	0.031
Denmark	0.009		0.009	Norway	0.008	0.394	0.006
Dominican Rep.	0.012		0.022	Pakistan	0.011		0.010
Ecuador	0.006		0.015	Portugal	0.007	0.495	0.014
Egypt	0.007	0.473	0.011	Seychelles	0.010		0.026
Finland	0.005	0.520	0.011	Sierra Leone	0.000		0.030
France	0.005	0.534	0.006	Spain	0.004	0.651	0.008
Greece	0.008	0.368	0.015	Sri Lanka	0.013		0.012
Guyana	0.003	0.309	0.023	Sudan	0.005		0.024
Honduras	0.004		0.013	Sweden	0.005	0.468	0.007
Hong Kong	0.021		0.018	Switzerland	0.004	0.328	0.009
Iceland	0.008	0.396	0.015	Syria	-0.265	0.009	0.032
Indonesia	0.011	0.324	0.017	United Kingdom	0.009		0.008
Ireland	0.009	0.465	0.010	Zimbabwe	-0.001	0.374	0.025
Italy	0.008	0.321	0.009				

Note: Table 1 gives Best Fitting Difference Stationary models for GDP series of various countries. Detail about this estimation of models is given in Section 2.3.

Table 2: Best Fitting Trend Stationary Models for Various Countries

Country	Const	Lag 1	Trend	Sigma	Country	Const	Lag 1	Trend	Sigma
Australia	0.700	0.83	0.001	0.008	Japan	0.257	0.95		0.010
Austria	0.119	0.97		0.007	Kenya	0.166	0.94		0.020
Belgium	0.135	0.97		0.007	Korea	0.502	0.84	0.004	0.015
Benin	0.311	0.87		0.013	Luxembourg	0.469	0.89	0.002	0.013
Botswana	0.070	0.99		0.019	Malawi	0.268	0.88		0.023
Burkina Faso	1.157	0.47	0.002	0.012	Malta	0.081	0.98		0.017
Burundi	0.222	0.89		0.023	Nepal	0.214	0.90	0.001	0.012
Cameroon	0.139	0.95		0.025	Netherlands	0.493	0.88	0.001	0.007
Central Africa	0.054	0.98		0.018	New Zealand	0.844	0.79	0.001	0.012
Chad	0.300	0.87		0.036	Nicaragua	0.077	0.97		0.031
China	0.445	0.74	0.008	0.026	Niger	0.670	0.74	-0.002	0.026
Cote d'Ivoire	0.297	0.90	-0.001	0.021	Nigeria	0.348	0.87		0.031
Denmark	1.001	0.76	0.002	0.008	Norway	0.200	0.96		0.007
Dominican Rep.	0.430	0.85	0.002	0.023	Pakistan	0.234	0.90	0.001	0.010
Ecuador	0.205	0.93		0.015	Portugal	0.271	0.93	0.001	0.014
Egypt	0.205	0.93	0.001	0.012	Seychelles	0.613	0.82	0.002	0.026
Finland	0.333	0.92	0.001	0.012	Sierra Leone	0.133	0.94		0.030
France	0.158	0.97	0.001	0.005	Spain	0.587	0.85	0.001	0.008
Greece	0.213	0.95		0.014	Sri Lanka	0.805	0.66	0.005	0.011
Guyana	0.115	0.96		0.023	Sudan	0.185	0.92	0.001	0.023
Honduras	0.438	0.85	0.001	0.012	Sweden	0.668	0.84	0.001	0.007
Hong Kong	0.114	0.98	0.001	0.017	Switzerland	0.784	0.82	0.001	0.008
Iceland	0.395	0.91	0.001	0.016	Syria	0.202	0.93		0.033
Indonesia	0.292	0.87	0.003	0.017	United Kingdom	0.989	0.76	0.002	0.007
Ireland	0.079	0.98	0.001	0.011	Zimbabwe	0.187	0.93		0.026
Italy	0.171	0.96		0.008					

Note: Table 2 gives Best Fitting Trend Stationary models for GDP series of various countries. Detail about this estimation is given in Section 2.3.

Table 3: Probability of Rejection of Unit Root (size) for Various DS Models

	Sigma=0.01						Sigma=0.02					
CNST	**DFN**	**DFC**	**DFT**	**PPN**	**PPC**	**PPT**	**DFN**	**DFC**	**DFT**	**PPN**	**PPC**	**PPT**
0	**4**	5	5	4	5	6	**4**	5	5	4	5	5
0.05	**0**	0	5	2	1	5	**1**	0	5	2	0	5
0.1	**0**	0	5	0	0	5	**0**	0	5	0	0	5
0.15	**0**	0	5	0	0	6	**0**	0	5	0	0	6
0.2	**0**	0	5	0	0	6	**0**	0	5	0	0	5
0.25	**0**	0	5	0	0	5	**0**	0	5	0	0	6

Note: Table 3 gives probability of rejection (size) of unit root tests for Difference Stationary models belonging to parametric space Q_{DS} . The variance does not seem to play any role in determining size of tests within the parametric space. The size of tests does not exceed nominal level of 5%.

Table 4: Probability of Rejection of Unit Root for Various TS Models (power)

CNST	LAG1	DFN	DFC	DFT	PPN	PPC	PPT	DFGC	DFGT	ZAC	ZTC	SBC	PTC	ZAT	ZTT	SBT	PTT
	0.86	7	32	13	12	39	16	13	12	11	11	11	5	4	4	4	4
	0.88	5	28	10	8	35	12	13	12	11	11	11	5	4	4	4	4
	0.9	3	24	8	6	29	10	10	9	8	8	8	5	3	3	4	3
	0.92	2	19	7	3	24	8	6	6	5	5	5	4	3	3	3	3
0.1	0.94	1	15	6	1	18	6	4	4	3	3	3	4	2	2	3	2
	0.96	1	10	5	1	12	5	1	1	1	1	1	4	3	3	3	2
	0.97	0	7	5	1	9	5	1	1	1	1	1	4	3	3	3	3
	0.98	0	5	5	0	6	6	0	0	0	0	0	5	3	3	3	3
	0.99	0	3	5	1	3	5	0	0	0	0	0	4	3	3	4	3
	0.86	0	48	18	1	58	23	7	7	6	6	6	3	2	2	2	2
	0.88	0	45	14	0	56	18	5	4	4	4	4	2	2	2	2	2
	0.9	0	44	11	0	54	13	3	3	2	2	2	2	1	1	1	1
	0.92	0	42	9	0	52	10	1	1	1	1	1	1	1	1	1	1
0.2	0.94	0	35	6	0	45	6	0	0	0	0	0	1	1	1	1	1
	0.96	0	25	4	0	33	5	0	0	0	0	0	1	1	1	1	1
	0.97	0	17	4	0	23	4	0	0	0	0	0	1	1	1	1	1
	0.98	0	11	4	0	14	4	0	0	0	0	0	2	2	1	2	2
	0.99	0	4	4	0	5	5	0	0	0	0	0	4	3	2	3	2
	0.86	0	67	26	0	78	36	2	2	2	2	2	1	1	1	1	1
	0.88	0	68	20	0	79	28	1	1	1	1	1	1	1	1	1	1
	0.9	0	69	15	0	80	21	0	0	0	0	0	0	0	0	0	0
	0.92	0	67	11	0	81	14	0	0	0	0	0	0	0	0	0	0
	0.94	0	65	7	0	78	9	0	0	0	0	0	0	0	0	0	0
0.3	0.96	0	53	4	0	68	5	0	0	0	0	0	0	0	0	0	0
	0.97	0	39	4	0	53	4	0	0	0	0	0	0	0	0	0	0
	0.98	0	24	4	0	33	4	0	0	0	0	0	1	1	0	1	0
	0.99	0	7	4	0	10	4	0	0	0	0	0	3	2	2	2	2
	0.86	0	84	37	0	91	53	1	1	0	0	0	0	0	0	0	0
	0.88	0	85	29	0	93	43	0	0	0	0	0	0	0	0	0	0
	0.9	0	87	23	0	94	33	0	0	0	0	0	0	0	0	0	0
0.4	0.92	0	87	15	0	96	22	0	0	0	0	0	0	0	0	0	0
	0.94	0	87	8	0	96	12	0	0	0	0	0	0	0	0	0	0
	0.96	0	79	4	0	92	5	0	0	0	0	0	0	0	0	0	0
	0.97	0	68	3	0	83	3	0	0	0	0	0	0	0	0	0	0

Table 4: Probability of Rejection of Unit Root for Various TS Models (power)

CNST	LAG1	DFN	DFC	DFT	PPN	PPC	PPT	DFGC	DFGT	ZAC	ZTC	SBC	PTC	ZAT	ZTT	SBT	PTT
	0.98	0	44	2	0	57	2	0	0	0	0	0	0	0	0	0	0
	0.99	0	14	3	0	19	3	0	0	0	0	0	1	1	1	1	1
	0.86	0	93	49	0	97	68	0	0	0	0	0	0	0	0	0	0
	0.88	0	95	43	0	98	61	0	0	0	0	0	0	0	0	0	0
	0.9	0	96	33	0	99	49	0	0	0	0	0	0	0	0	0	0
	0.92	0	97	22	0	99	34	0	0	0	0	0	0	0	0	0	0
0.5	0.94	0	97	11	0	100	17	0	0	0	0	0	0	0	0	0	0
	0.96	0	93	5	0	99	6	0	0	0	0	0	0	0	0	0	0
	0.97	0	86	3	0	96	3	0	0	0	0	0	0	0	0	0	0
	0.98	0	67	2	0	81	2	0	0	0	0	0	0	0	0	0	0
	0.99	0	22	3	0	32	3	0	0	0	0	0	1	0	0	1	0

Note: Table 4 gives probability of rejection of unit root tests for trend stationary models (power) belonging to parametric space Q_{TS} . Simulation was done for different values of variance but the variance did not affect power of tests.

Table 5: Power of PPC Test for Various Countries

Country	Estimated TS Model		Power	Country	Estimated TS Model		Power
	Const	Lag			Const	Lag	
Malta	0.081	0.98	5	Greece	0.213	0.95	36
Nicaragua	0.077	0.97	7	Zimbabwe	0.187	0.93	46
Austria	0.119	0.97	10	Japan	0.257	0.95	48
Belgium	0.135	0.97	11	Syria	0.202	0.93	51
Guyana	0.115	0.96	13	Ecuador	0.205	0.93	52
Italy	0.171	0.96	21	Burundi	0.222	0.89	78
Cameroon	0.139	0.95	21	Chad	0.3	0.87	89
Norway	0.2	0.96	26	Malawi	0.268	0.88	89
Sierra Leone	0.133	0.94	26	Benin	0.311	0.87	91
Kenya	0.166	0.94	33	Nigeria	0.348	0.87	98

Citation: Rehman, A. (2019). Detecting stationarity of GDP: A test of unit root tests, *Journal of Quantitative Methods*, *3*(1), 8-37.

Submission Date: May 01, 2018
Last Revised: February 07, 2019
Acceptance Date: February 27, 2019