Review Article

Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order: Critical Comments

Hari S. Luitel[1] • **Gerry J. Mahar**[2]

Abstract In this research note, we accomplish two objectives. First, we reexamine the reliability of unit root findings in the study by Said and Dickey (1984) and show that their results are internally consistent. Second, we provide new results from the reanalysis of the original data that were not included in their study and explain why their results cannot be generalized. In conclusion, we cast doubt on the continued usefulness of Augmented Dickey Fuller (ADF) test as a sound scientific method.

Key Words: Econometrics; Statistics; Methodology; History of Thought

JEL Classification: C10, B00

Introduction

Statistical theory relating to the first order autoregressive unit root process where the autoregressive parameter is equal to one (unstable process) and greater than one (explosive process) dates back to early writings of Mann and Wald (1943), Rubin (1950), Anderson (1959), White (1958, 1959), and

1. *hari.luitel@algomau.ca*

2. *gerry.mahar@algomau.ca*

Rao (1961). However, Maddala and Kim (1998, p. 3) together with Banerjee et. all (1993, p. 1), Phillips (1995) and Patterson (2011) leave one with the impression that the seminal work of Nelson and Plosser (1982) marked a paradigm shift in the macroeconomics literature in the 1980s. In the 1960s and 1970s, conventional practice in time series analysis was to work with data that were differenced a sufficient number of times to render them stationary.[3] This practice was based on informal diagnostics rather than formal statistical tests and Nelson and Plosser (1982) replaced the practice with formal Dickey-Fuller (DF) tests for unit roots (Stock 1994, p. 2741). For a positive review of history of unit root tests, see Fuller (1984), Stock (1994) and Patterson (2011, 2012). For surveys on specialized topics in unit roots such as changing the mean or available methods to choose for a length of time lag etc., see Diebold and Nerlove (1990), Perron (1990), Campbell and Perron (1991), and Ng and Perron (2001). A further review of the literature on unit root in conjunction with structural breaks can be found in Stock (1994) and in Perron and Zhu (2005). The implications of unit root in economic theory, policy and econometric procedures, were discussed in Chinn (1991) and Libanio (2005).

The literature on unit roots remains controversial because many tests results of unit root rest on a razor's edge. Although the literature is extensive and often studies failed to reject the hypothesis $\rho = 1$, many also failed to reject the hypothesis $\rho \geq 0.95$ at 95 percent, or even 99 percent, confidence interval level (Greene 2000, p. 781). An early critique by Cochrane (1991), Maddala (1992, p.582-588), Harvey (1997), Maddala and Kim (1998), Phillips (2003) and a recent critique by Moosa (2011), Luitel and Mahar (2015a, 2015b, 2016), and Luitel et. al. (2018) further highlight the continuing controversy surrounding the unit root test in the literature.

Our focus in this paper will be on the Dickey-Fuller (DF) and the Augmented Dickey-Fuller (ADF) tests of unit root. These tests are used routinely to decide whether a time series would be stationary or nonstationary. Because these tests predate all tests for nonstationarity in a time series, we argue that these

3 A time series is said to be stationary if both its marginal and all joint distributions are independent of time.

tests likely led to an early widespread use resulting in publication bias.[4] We also acknowledge that unit root tests are now a standard diagnostic tool in applied time series analysis. However, we find it very curious that leading econometrics textbooks such as Hamilton (1994), Greene (2000), Hayashi (2000), Davidson and Mackinnon (2004), and Patterson (2011, 2012) among others do not contain explicit warnings against the dangers of the use of unit root tests. It is possible that earlier warnings may have faded away or may have been weakly stated because they appeared infrequently in the form of journal articles. Since the well-respected applied works still make frequent use of unit root tests, it is timely for a critical review of this method.[5]

We have two objectives in this paper: First, we reexamine the reliability of unit root findings in the study by Said and Dickey (1984).[6] We can reaffirm the internal validity of their research design. Second, we found new results not included in that study and explain why the new results cast doubt on the usefulness of the ADF test to determine a unit root in a time series.

Critical Review of Dickey-Fuller and Augmented Dickey-Fuller Tests

Researchers working on econometric time series analysis acknowledge that the literature on unit root is vast and often confusing (Phillips and Xiao 1998; Bierens 2001; Patterson 2011). From an applied practitioner's viewpoint, it remains unclear which tests, if any, are superior to others and much empirical work continues to use the simple testing procedures, the DF and the ADF tests (Phillips and Xiao 1998). Therefore, in this section, we limit our discussion of tests for unit root to a critical review of the DF and the ADF tests.[7]

4 Publication bias arises if statistically significant results are more likely to be published than other results (Miguel, 2015, p. 5). As such, over time the published literature becomes systematically unrepresentative of the population of completed studies. For a detailed discussion of nature, sources, consequences as well as remedies of publication bias, see Rothstein, Sutton and Borenstein (2005).

5 The unit root tests have made inroads into other disciplines beyond economics. For example, research work into environmental science, such as the analysis of climate change or paleoclimate data, see Kaufmann et al. (2006, 2013), Davidson et al. (2016), Storelvmo et al. (2016).

6 On the topic of testing for unit roots, the 1984 paper by Said and Dickey was the 20th most influential paper out of top 100 highly influential papers published by *Biometrika* since 1936 (Titterington, 2013, p. 34). See also Patterson (2011, preface).

7 The anomalies that arise from the use of panel unit root tests are discussed in Luitel and Mahar (2021).

Pure Random Walk Model

Many studies of the asymptotic behavior of the OLS estimator $\hat{\rho}$ begin with an AR(1) model without a deterministic trend. Dickey and Fuller (1979, p. 427) considered the following regression equation:

$$Y_t = \rho Y_{t-1} + e_t\,,\ t = 1, 2, \ldots \tag{1}$$

where Y_0=0, ρ was a real number, and $\{e_t\}$ was a sequence of independent normally distributed random variables with a mean zero and variance σ^2 [i.e., e_t NID(0, σ^2)]. ρ was obtained using the ordinary least square (OLS) method:

$$\hat{\rho} = \left(\sum_{t=1}^{n} Y_{t-1}^{2}\right)^{-1} \sum_{t=1}^{n} Y_t Y_{t-1}\,.$$

To determine the coefficient of $\hat{\rho}$ above, Dickey and Fuller (1979, p.427) proposed $\hat{\tau} = (\hat{\rho} - 1)\, S_e^{-1} \left(\sum_{i=2}^{n} Y_{t-1}^{2}\right)^{\frac{1}{2}}$ where $S_e^{2} = (n-2)^{-1} \sum_{i=2}^{n} (Y_t - \hat{\rho}\, Y_{t-1})^2$. For applied economic research, the hypothesis that ρ - 1 will be of particular interest because the least square method will be invariant to a linear transformation and the hypothesis that ρ - 1 enables researchers to transform data by adding, subtracting or doing other manipulations. For example, by subtracting Y_{t-1} from both sides, equation (1) may be written as:

$$\Delta Y_t = \delta Y_{t-1} + e_t \tag{2}$$

where $\delta = \rho - 1$. We recognize the practice of transforming a time series in this manner as differencing. A time series Y_t is referred to as being a difference stationary if differencing converts Y_t into a stationary time series. In general, a difference stationary process is a process whereby a time series can be converted into stationarity by differencing. It was in this sense that equation (1), or its linear transformation as shown in equation (2), also became known as a pure random walk model.

Random Walk with Drift

For many applications to economic data, a pure random walk model may be too restrictive. For example, when applied to the relationship between disposable income and consumer consumption as in permanent income hypothesis (Friedman, 1957) or the life cycle model of consumption (Modigliani, 1986). Even though income was nil at a point in time, it did not necessarily mean that consumer consumption was nil at that point in time. There are other examples and the implication was that not all regression equations would have a zero intercept term. For the pure random walk model, this implication was accommodated by the introduction of a drift term such as:

$$Y_t = \alpha + \rho Y_{t-1} + e_t \tag{3}$$

where ρ = 1 corresponded to a unit root. In equation (3), the intercept was argued to satisfy α = (1- ρ)μ, where μ was the mean of the time series and the null hypothesis of a unit root implied that the intercept term should be zero. In principle, it was possible to jointly test the two restrictions α = 0 and ρ = 1 in equation (3). In practice, however, a more convenient method, equation (4), was used to test the null hypothesis ρ = 1. Essentially, equation (4) was a linear transformation of equation (3) and was obtained by subtracting Y_{t-1} from both sides of equation (3):

$$\Delta Y_t = \alpha + \delta Y_{t-1} + e_t \tag{4}$$

where $\delta = \rho$ -1. If $\alpha \neq 0$ and $\rho = 1$, then equation (3), or its linear transformation equation (4), was known as a random walk with drift, with α being the drift parameter. For the level variable Y_t, α corresponded to a linear time trend. If two conditions $\alpha \neq 0$ (a nonzero intercept) and ρ - 1 (unit root) hold, then α did not equal (1- ρ)μ, in which case, equation (4) could not be derived from a pure AR(1) model. This can be seen by considering the resulting process:

$$\Delta Y_t = \alpha + e_t \tag{5}$$

Taking expectations of equation (5), E{ΔY_t} = α, and for a given starting value Y_0, E{ΔY_t} = $Y_0 + \alpha t$. Thus, the interpretation of the intercept term in equation (4) depended on whether or not there was a unit root. In the stationary case, α was interpreted to be the non-zero mean of the series, whereas in the nonstationary case (unit root), it was interpreted as a deterministic trend in Y_t.

Random Walk with Drift and Time Trend

Another argument put forward in the development of a test for unit root was that nonstationarity in a time series might arise due to the presence of a deterministic time trend. For a time series that had clear time trends, the AR(1) model was modified as:

$$Y_t = \alpha + \beta \mathrm{t} + \rho Y_{t-1} + e_t \tag{6}$$

with $|\rho|<1$ and $\beta \neq 0$. Equation (6) showed that a nonstationary process due to a linear trend, βt, could be removed by regressing Y_t on a constant and t, and then considering the residuals of this regression, or by including t as an additional variable in the model. This process for Y_t was referred to as being trend stationary.

It was possible to test the hypothesis whether Y_t followed a random walk against the alternative hypothesis that it followed a trend stationary process as in equation (6). However, equation (7), a linear transformation of equation (6) by subtracting Y_{t-1} from both sides, was more commonly used:

$$\Delta Y_t = \alpha + \beta \mathrm{t} + \delta Y_{t-1} + e_t \tag{7}$$

where $\delta = \rho$ -1. In fact, equation (7) provided the basis for all three different DF tests for the existence of unit roots. A test of the hypothesis that α, β and δ equal zero confirmed the pure random walk model. A test of the hypothesis that β and δ equal zero confirmed the model of random walk with drift. If δ was less than zero, then the evidence favored the trend stationary model, and detrending was considered an appropriate approach.

The parameter of interest in the regression equation (7) was the value of δ, yet the distribution of test statistics depended crucially on the nuisance parameters of the α and β coefficients. Under the null hypothesis $\delta = 0$, the standard t-ratio did not have a Student's t distribution, not even asymptotically. Although the null hypothesis would be the same in all three cases, the testing regression would be different and there would be a different distribution of the test statistics. For various combinations of the three data generating processes, Dickey and Fuller (1979, 1981) proposed three different asymptotic distributions of test statistics that corresponded to the model used to test the null hypothesis for the existence of a unit root. These test statistics were updated later by Guilkey and Schmidt (1989), Schmidt (1990) and Mackinnon (1991, 1994).[8] Interestingly, the critical values of the DF_{τ} were systematically smaller than those for *DF*.

Autoregressive Moving Average of Unknown Order

In our review of the literature, we have identified two reasons why the Dickey-Fuller tests described above were not satisfactory. First, the DF unit root test regressions, equation (2), equation (4) and equation (7), were based on the assumption that the data generating process was an AR(1) process and thus did not include any lagged variable beyond Y_{t-1} . Bierens (2001, p. 620) argued that this assumption was not very realistic because even after differencing, most macroeconomic time series would likely display a fair amount of dependence. Second, the error terms e_t in equations (2), (4) and (7) may also be serially correlated. If the error terms e_t were serially correlated, then the Dickey-Fuller tests would no longer be asymptotically valid (Davidson and Mackinnon 2004, p.620, Hayashi 2000, p. 585). *We acknowledge these arguments for our research purpose in this paper.*

To overcome these criticisms, several studies recommended a modification to the DF test by adding an appropriate number of lagged differences to the autoregressive equation. For example, Harris (1992) proposed a formula, $l_i = int\{i(n/100)^{1/4}\}$ to determine the lag length that allowed for the order

8 For practical purpose, many econometric computer software routinely report interpolated DF test statistics at 90 percent, 95 percent and 99 percent confidence level.

of autoregression to grow with sample size (page 383). Previously, Schwert (1989, page 151) used lag lengths based on the formulas $l_4 = int\{4(T/100)^{1/4}\}$ and $l_{12} = int\{12(T/100)^{1/4}\}$. In contrast, Taylor (2000) recommended selecting a lag length using a data-based algorithm or by using a much higher level of significance (*e.g. 0.2 level rather than the traditional 0.05 level*) in the general-to-specific rule framework. The practice of adding an appropriate number of lagged differences to the autoregressive equation became known later as the Augmented Dickey-Fuller (ADF) test.

Consider the following AR(p) process:

$$Y_t = \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + \dots \rho_p Y_{t-p} + e_t \tag{8}$$

To obtain a linear transformation by subtracting Y_{t-1} from both sides, proponents of unit root tests argue that equation (8) could be written as:

$$\Delta Y_t = \delta Y_{t-1} + \sum_{i=1}^{p} \gamma_i Y_{t-i} + e_t \tag{9}$$

where $\delta = \rho_1 + \rho_2 + \dots + \rho_p - 1$. Equation (9) became known as the Augmented Dickey-Fuller (ADF) regression. Under the null hypothesis, $\delta = 0$, ΔY_t became a stationary AR(p) process, while under the alternate hypothesis, $\delta < 0$, Y_t (original series) became a stationary process.

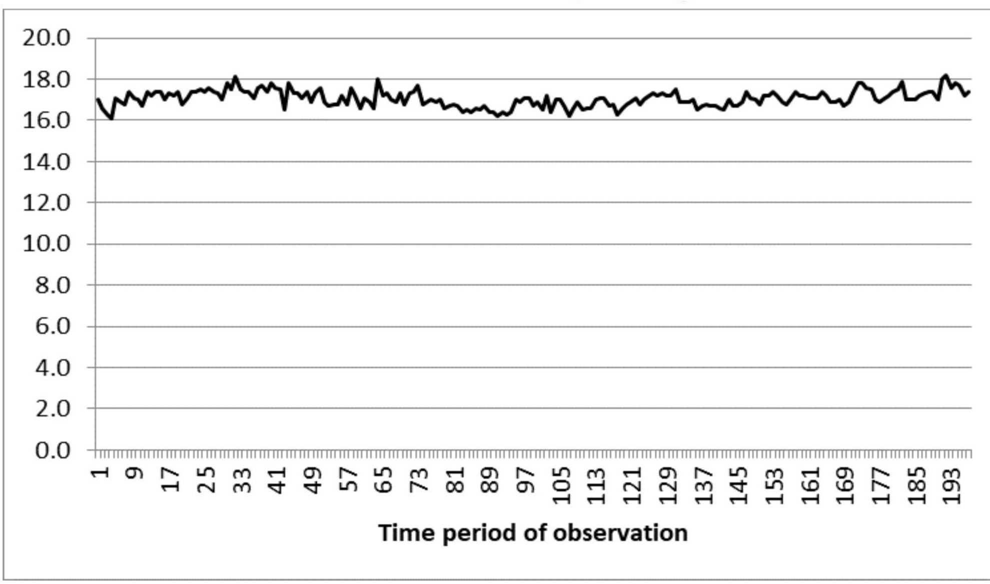
The source of the controversy with the ADF test centered around the arbitrary choice of the length of time lag. Although the *Akaike* or *Schwarz information criteria* or some other rules were generally used to decide the length of time lag, they did not always yield unique results. We revisited the source of this controversy and found that none of the studies in the extant literature went far enough to point out the limits of the DF and the ADF tests. It is this gap in the literature that we attempt to fill in this paper. Particularly, we report new statistical results from the reanalysis of the original data that were not included in the study by Said and Dickey (1984) and show the sensitivity of the length of time lag in their results. We then explain why the new results cast doubt on the continued usefulness of the ADF test.

Data Analysis and Results

In this section, we perform post-publication replication of the data previously analyzed by Said and Dickey (1984) and report that our results go further than a mere replication exercise.[9] Said and Dickey (1984, p. 606-607) used secondary data that was originally reported by Box and Jenkins (1970, p. 525) concerning the concentration readings arising from a chemical process.[10] We denote this chemical process as Y_t and Figure 1 shows the actual observation of Y_t. As can be observed in Figure 1, Y_t varies between a narrow range of 16.1 and 18.2 with a mean value of 17.0624. Said and Dickey (1984) reported this time series as nonstationary. We are curious to reanalyze the original data to ascertain whether the series is in fact nonstationary or stationary.

Figure 1

"Uncontrolled" Concentration, Two-Hourly Readings: Chemical Process



Source: Box and Jenkins (1970, p. 525)

9 For a surge of interest in replication of published research in economics, see Hamermesh (2007), McCullough, McGeary and Harrison (2008), McCullough (2009), McCullough and McKitrick (2009), Burman, Reed and Alm (2010), Chang and Li (2015), Duvendack, Palmer-Jones and Reed (2015), Zimmerman (2015). For various aspects of replication and their implications on research outcome, see Clemens (2017).

10 For further information how the data was collected, see Box and Jenkins (1970, p.85).

Table 1 presents the summary statistics of Y_t and all other auxiliary variables derived from Y_t that were involved in the analysis. It may be unconventional to report summary statistics of the auxiliary variables obtained from a time series but we break with this convention and report them to learn if they reveal additional information that the summary statistics of Y_t alone did not provide. As seen in Table 1, except for Y_t and Y_{t-1} the mean values of all variables are 0 up to a hundredth decimal place. The most striking feature of these auxiliary variables, however, is that all variables have a minimum value of -1 and a maximum value of +1.4.

Table 1 Summary Statistics

Variables	Number of observations	Mean	Standard Deviation	Minimum	Maximum
$Y_{t-1} - \bar{Y}$	196	-0.0017	0.3995	-0.9624	1.1375
Y_t	197	17.0624	0.3992	16.1	18.2
Y_{t-1}	196	17.0607	0.3995	16.1	18.2
$\dot{Y}_t$	196	0.0020	0.3703	-1	1.4
$\dot{Y}_{t-1}$	195	0.0010	0.3709	-1	1.4
$\dot{Y}_{t-2}$	194	0.0036	0.3701	-1	1.4
$\dot{Y}_{t-3}$	193	0.0041	0.3710	-1	1.4
$\dot{Y}_{t-4}$	192	0.0031	0.3717	-1	1.4
$\dot{Y}_{t-5}$	191	0.0062	0.3701	-1	1.4
$\dot{Y}_{t-6}$	190	0.0052	0.3708	-1	1.4
$\dot{Y}_{t-7}$	189	0	0.3646	-1	1.4
$\dot{Y}_{t-8}$	188	0.0021	0.3644	-1	1.4
$\dot{Y}_{t-9}$	187	0.0021	0.3654	-1	1.4
$\dot{Y}_{t-10}$	186	0.0016	0.3663	-1	1.4

Source: Box and Jenkins (1970, p. 525)

Said and Dickey argued that an autoregressive model of order 10 would give a sufficient approximation to the data. Although this argument may have followed from Theorem § 6 of Berk (1974, p. 501), the limit theory does not specify the exact value of the length of time lag for any given number of observation. Said and Dickey therefore performed sensitivity of their results with the lagged differences 6, 7, 8, 9 and 10. For reporting final results, however, Said and Dickey (1984) fitted the following regression model:

$$\dot{Y}_t = \delta(Y_{t-1} - \bar{Y}) + \sum_{i=1}^{6} \gamma_i \dot{Y}_{t-i} \quad (10)$$

where $\dot{Y}_t = Y_t - Y_{t-1}$. Note that the regression equation (10) did not have a constant term and it included lagged differences 1, 2, 3 4, 5 and 6.

The replication results are reported in Table 2. In the table, column 1 presents regression results from equation (10). From these results, the Studentized statistic can be calculated as $(0.0785)^{-1}$ (-0.1601) = -2.04. Comparing the $\hat{\tau}$ tables of Fuller (1976, p. 373), Said and Dickey (1984) did not reject the null hypothesis of unit root using a one sided tail at 10 percent significance level. Although Said and Dickey reported separately the DF $\hat{\tau}$ statistics for lagged differences 7, 8, 9 and 10, they did not report the regression coefficients. We report both in Table 2, column 2, column 3, column 4, and column 5. When reporting the DF $\hat{\tau}$ statistics for lagged differences 7, 8, 9 and 10, however, we discovered that Said and Dickey switched to the regression model with a constant term, as opposed to the regression equation (10) noted above.

Using the procedure described in Said and Dickey's paper and independent of the original authors, we obtained identical results. Thus, we successfully replicated their findings and confirm that the results reported by Said and Dickey were internally valid. This provided us with confidence that we were moving in the right direction in our replication exercise. We are now in a position to explain why these results are unable to overcome the threats to external validity, which we present in the next section.[11]

Extensions and Robustness Checks

In this section, we perform several extensions and robustness checks to examine the sensitivity of the results. Table 3 and Table 4 present additional results that Said and Dickey did not include in their study. Table 3 presents results when constant term was not included in the regression equation for

11 External validity refers to generalizability of statistical inferences and conclusions based on one population and setting being studied to other populations and settings. On the other hand, internal validity refers to the statistical inferences and conclusions that are valid for the population and setting being studied (Stock and Watson, 2003).

Table 2 Said and Dickey (1984) Replication Results

	No constant	Constant term included			
$\dot{Y}_t$	Column (1)	Column (2)	Column (3)	Column (4)	Column (5)
$Y_{t-1} - \bar{Y}$	-0.1601** (0.0785)	-	-	-	-
Y_{t-1}	-	-0.1546* (0.0800)	-0.1501* (0.0820)	-0.1510* (0.0841)	-0.1725** (0.0856)
$\dot{Y}_{t-1}$	-0.4941*** (0.0963)	-0.4901*** (0.1005)	-0.4975*** (0.1028)	-0.4969*** (0.1049)	-0.4756*** (0.1062)
$\dot{Y}_{t-2}$	-0.2919*** (0.0985)	-0.2907*** (0.1041)	-0.2984*** (0.1087)	-0.3005*** (0.1114)	-0.2770** (0.1128)
$\dot{Y}_{t-3}$	-0.2640*** (0.0947	-0.2803*** (0.1015)	-0.2889*** (0.1076)	-0.2872** (0.1124)	-0.2611** (0.1144)
$\dot{Y}_{t-4}$	-0.2477*** (0.0903)	-0.2405** (0.0971)	-0.2525** (0.1045)	-0.2501** (0.1107)	-0.2099* (0.1148)
$\dot{Y}_{t-5}$	-0.2681*** (0.0858)	-0.2486*** (0.0929)	-0.2563** (0.1000)	-0.2563** (0.1076)	-0.2105* (0.1131)
$\dot{Y}_{t-6}$	-0.1888*** (0.0726)	-0.1593* (0.0882)	-0.1641* (0.0952)	-0.1602 (0.1023)	-0.1082 (0.1094)
$\dot{Y}_{t-7}$	-	0.0346 (0.0749)	0.0283 (0.0901)	0.0341 (0.0971)	0.0808 (0.1037)
$\dot{Y}_{t-8}$	-	-	-0.0119 (0.0754)	-0.0034 (0.0907)	.0362 (.0971)
$\dot{Y}_{t-9}$	-	-	-	0.0115 (0.0758)	0.0677 (0.0907)
$\dot{Y}_{t-10}$	-	-	-	-	0.0899 (0.0759)
Constant	-	2.6486* (1.3656)	2.5730* (1.4000)	2.5875* (1.4348)	2.9531** (1.4614)
DF $\hat{\tau}_\mu$ Statistics	-2.04	-1.931	-1.830	-1.796	-2.013
Error Mean Square	0.0937	0.0935	0.0945	0.0955	.0956
Error Sum of Square	17.1591	16.8397	16.8248	16.8168	16.6423
F statistics	F(7, 183) = 12.56***	F(8, 180) = 10.96***	F(9, 178) = 9.55***	F(10, 176) = 8.50***	F(11, 174) = 7.80***
R^2	0.3244	0.3275	0.3256	0.3257	0.3302
Adjusted R^2	0.2986	0.2976	0.2915	0.2874	0.2879
Root MSE	0.3062	0.3058	0.3074	0.3091	0.3092
No of Obser-vations	190	189	188	187	186
Main Finding	Nonstationary	Nonstationary	Nonstationary	Nonstationary	Nonstationary

Notes: Figures in parenthesis are standard errors. ***indicates 1 percent significance level, **indicates 5 percent significance level and * indicates 10 percent significance level.

Source: Box and Jenkins (1970, p. 525)

Table 3: Additional Results No Constant Term Included

$\dot{Y}_t$	Column (1)	Column (2)	Column (3)	Column (4)	Column (5)	Column (6)
Y_{t-1} - Ȳ	-0.4277*** (0.0588)	-0.3198*** (0.0643)	-0.3002*** (0.0681)	-0.2852*** (0.0714)	-0.2484*** (0.0743)	-0.2062*** (0.0770)
$\dot{Y}_{t-1}$	-	-0.2534*** (0.0694)	-0.3038*** (0.0804)	-0.3412*** (0.0845)	-0.3665*** (0.0880)	-0.4214*** (0.0926)
$\dot{Y}_{t-2}$	-	-	-0.0785 (0.0718)	-0.1395* (0.0837)	-0.1578* (0.0884)	-0.2195** (0.0932)
$\dot{Y}_{t-3}$	-	-	-	-0.0852 (0.0719)	-0.1164 (0.0841)	-0.1787** (0.0897)
$\dot{Y}_{t-4}$	-	-	-	-	-0.0694 (0.0717)	-0.1575* (0.0846)
$\dot{Y}_{t-5}$	-	-	-	-	-	-0.1437** (0.0726)
$\dot{Y}_{t-6}$	-	-	-	-	-	-
$\dot{Y}_{t-7}$	-	-	-	-	-	-
$\dot{Y}_{t-8}$	-	-	-	-	-	-
$\dot{Y}_{t-9}$	-	-	-	-	-	-
$\dot{Y}_{t-10}$	-	-	-	-	-	-
DF $\hat{\tau}_\mu$ Statistics	-7.27***	-4.97***	-4.39***	-3.98***	-3.30***	-2.61***
Error Mean Square	0.1079	0.1011	0.0999	.0986	.0973	0.0962
Error Sum of Square	21.0445	19.5151	19.0912	18.6408	18.1977	17.8129
F statistics	F(1, 195) = 52.77***	F(2, 193) = 34.93***	F(3, 191) = 24.67***	F(4, 189) = 19.79***	F(5, 187) = 14.90***	F(6, 185) = 13.15***
R^2	0.2130	0.2658	0.2793	0.2952	0.2850	0.2990
Adjusted R^2	0.2090	0.2582	0.2680	0.2803	0.2658	0.2762
Root MSE	0.3285	0.3179	0.3161	0.3140	0.3119	0.3103
No of Observations	196	195	194	193	192	191
Main Finding	Stationary	Stationary	Stationary	Stationary	Stationary	Stationary

Notes: Figures in parenthesis are standard errors. ***indicates 1 percent significance level, **indicates 5 percent significance level and * indicates 10 percent significance level.

Source: Box and Jenkins (1970, p. 525)

Table 3: Additional Results No Constant Term Included (continued…)

$\dot{Y}_t$	Column (7)	Column (8)	Column (9)	Column (10)
$Y_{t-1} - \bar{Y}$	-0.1554*	-0.1512*	-0.1522*	-0.1738**
	(0.0798)	(0.0818)	(.0838)	(0.0854)
$\dot{Y}_{t-1}$	-0.4877***	-0.4953***	-0.4947***	-0.4733***
	(0.1001)	(0.1025)	(0.1046)	(0.1058)
$\dot{Y}_{t-2}$	-0.2874***	-0.2949***	-0.2973***	-0.2739**
	(0.1037)	(0.1082)	(0.1109)	(0.1122)
$\dot{Y}_{t-3}$	-0.2768***	-0.2850***	-0.2831**	-0.2574**
	(0.1010)	(0.1070)	(0.1117)	(0.1138)
$\dot{Y}_{t-4}$	-0.2366**	-0.2485**	-0.2457**	-0.2055*
	(0.0965)	(0.1039)	(0.1100)	(0.1140)
$\dot{Y}_{t-5}$	-0.2446***	-0.2517**	-0.2518**	-0.2058*
	(0.0923)	(0.0993)	(0.1069)	(0.1122)
$\dot{Y}_{t-6}$	-0.1559*	-0.1600*	-0.1556	-0.1037
	(0.0878)	(0.0946)	(0.1016)	(0.1086)
$\dot{Y}_{t-7}$	.0365	0.0314	0.0379	0.0850
	(0.0747)	(0.0897)	(0.0964)	(0.1030)
$\dot{Y}_{t-8}$	-	-0.0101	-0.0004	0.0398
		(0.0751)	(0.0902)	(0.0965)
$\dot{Y}_{t-9}$	-	-	0.0132	0.0705
			(0.0756)	(0.0902)
$\dot{Y}_{t-10}$	-	-	-	0.0915
				(0.0756)
DF $\hat{\tau}_\mu$ Statistics	-1.96	-1.86	-1.82	-2.03
Error Mean Square	0.0931	.0941	.0951	.0951
Error Sum of Square	16.8614	16.8450	16.8349	16.6586
F statistics	F(8, 181) = 10.97***	F(9, 179) = 9.57***	F(10, 177) = 8.52***	F(11, 175) = 7.82***
R^2	0.3266	0.3248	0.3250	0.3296
Adjusted R^2	0.2969	0.2909	0.2868	0.2875
Root MSE	0.3052	0.3067	0.3084	0.3085
No of Observations	189	188	187	186
Main Finding	Nonstationary	Nonstationary	Nonstationary	Nonstationary

Notes: Figures in parenthesis are standard errors. ***indicates 1 percent significance level, **indicates 5 percent significance level and * indicates 10 percent significance level.

Source: Box and Jenkins (1970, p. 525) lag differences 1 through 10 (except for lag differences 6), whereas Table 4 presents results when constant term was included in the regression equation for lag differences 1 through 6. In essence, Table 3 and Table 4 complement Table 2.

Out of possible 22 regressions results reported in the tables, results from 10 regression models (Table 2: column 1, column 2, column 3, column 4, column 5; Table 3: column 7, column 8, column 9, column 10; and Table 4: column 7) indicate that the series was nonstationary. In contrast, results from 12 regression models (Table 3: column 1, column 2, column 3, column 4, column 5, column 6; Table 4: column 1, column 2, column 3, column 4, column 5 and column 6) indicate that the series was stationary. In other words, more than 50 percent of the unit root test results indicated that the series was stationary and yet Said and Dickey reported that it was nonstationary. Given the controversial nature of unit root literature, these results suggest that Said and Dickey selectively reported only part of the results showing a confirmation bias in support of Box and Jenkins' (1970; p. 94) results.

Our primary interest to reanalyze the original data was to find out whether that time series was nonstationary or stationary. Clearly, the answer depends on whether one includes more than or less than 5 lagged differences in the data analysis. These results partly explain why and/or how the ADF test allows researchers to pick and choose the length of time lag to obtain desired results.

As noted in the final argument of the development of ADF test in Section 2.4, the extant literature recommended modifying the test by adding an appropriate number of lagged differences to the autoregressive equation using either information criteria or data-based algorithm. One of the two reasons these recommendations were made was that it would take into account the shortcoming of most macroeconomic time series that showed a fair amount of dependence even after differencing. We were therefore curious to examine the correlation coefficients of the auxiliary variables whether they revealed any dependence after first differencing. The results of correlation coefficients are reported in Table 5 and we found that the correlation coefficients were identical between Y_{t-1} - $\bar{Y}$ and other variables and between Y_{t-1} and other variables. However, contrary to our expectations from the reading of the literature, the correlation coefficients declined dramatically after the lagged values of the first

Table 4: Additional Results with Constant Term Included

$\dot{Y}_t$	Column (1)	Column (2)	Column (3)	Column (4)	Column (5)	Column (6)	Column (7)
Y_{t-1}	-0.4277*** (0.0590)	-0.3197*** (0.0644)	-0.3001*** (0.0683)	-0.2849*** (0.0716)	-0.2483*** (0.0745)	-0.2058*** (0.0771)	-0.1591** (0.0786)
$\dot{Y}_{t-1}$	-	-0.2535*** (0.0696)	-0.3042*** (0.0806)	-0.3423*** (0.0847)	-0.3676*** (0.0883)	-0.4226*** (0.0929)	-0.4965*** (0.0965)
$\dot{Y}_{t-2}$	-	-	-0.0789 (0.0720)	-0.1408* (0.0839)	-0.1593* (0.0887)	-0.2217** (0.0936)	-0.2953*** (0.0988)
$\dot{Y}_{t-3}$	-	-	-	-0.0861 (0.0720)	-0.1177 (0.0844)	-0.1811** (0.0900)	-0.2683*** (0.0951)
$\dot{Y}_{t-4}$	-	-	-	-	-0.0701 (0.0719)	-0.1595* (0.0849)	-0.2519*** (0.0906)
$\dot{Y}_{t-5}$	-	-	-	-	-	-0.1452** (0.0728)	-0.2720*** (0.0861)
$\dot{Y}_{t-6}$	-	-	-	-	-	-	-0.1913*** (0.0728)
Constant	7.2994*** (1.0074)	5.4597*** (1.1003)	5.1276*** (1.1657)	4.8733*** (1.2222)	4.2454*** (1.2720)	3.5221*** (1.3170)	2.7294** (1.3423)
DF $\hat{\tau}_\mu$ Statistics	-7.246***	-4.959***	-4.393***	-3.979***	-3.331**	-2.667*	-2.022
Error Mean Square	0.1084	0.1016	0.1004	0.0990	0.0977	0.0967	0.0940
Error Sum of Square	21.0442	19.5122	19.0812	18.6179	18.1846	17.7943	17.1222
F statistics	F(1, 194) = 52.50***	F(2, 192) = 34.76***	F(3, 190) = 24.57***	F(4, 188) = 19.75***	F(5, 186) = 14.86***	F(6, 184) = 13.12***	F(7, 182) = 12.57***
R^2	0.2130	0.2658	0.2795	0.2959	0.2855	0.2997	0.3258
Adjusted R^2	0.2089	0.2582	0.2681	0.2809	0.2663	0.2768	0.2999
Root MSE	0.3293	0.3187	0.3169	0.3146	0.3126	0.3109	0.3067
No of Observations	196	195	194	193	192	191	190
Main Finding	Stationary	Stationary	Stationary	Stationary	Stationary	Stationary	Nonstationary

Notes: Figures in parenthesis are standard errors. ***indicates 1 percent significance level, **indicates 5 percent significance level and * indicates 10 percent significance level.

Source: Box and Jenkins (1970, p. 525) differences. These results run counter to the justification for the practice of adding additional number of lagged differences to the autoregressive equation on the grounds that most macroeconomic time series display a fair amount of dependence even after differencing.

Table 5: Matrix of Correlation Coefficients

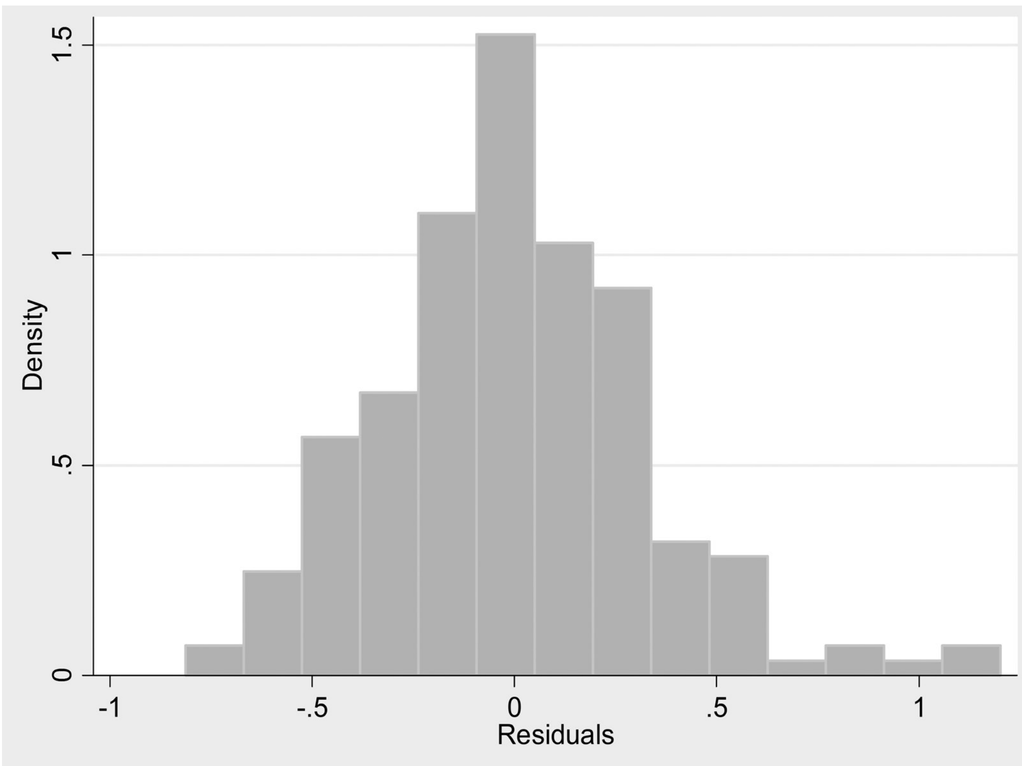
	$Y_{t-1} - \bar{Y}$	Y_t	Y_{t-1}	$\dot{Y}_t$	$\dot{Y}_{t-1}$	$\dot{Y}_{t-2}$	$\dot{Y}_{t-3}$	$\dot{Y}_{t-4}$	$\dot{Y}_{t-5}$	$\dot{Y}_{t-6}$	$\dot{Y}_{t-7}$	$\dot{Y}_{t-8}$	$\dot{Y}_{t-9}$	$\dot{Y}_{t-10}$
$Y_{t-1} - \bar{Y}$	1.0000													
Y_t	0.5738	1.0000												
Y_{t-1}	1.0000	0.5738	1.0000											
$\dot{Y}_t$	-0.4622	0.4611	-0.4622	1.0000										
$\dot{Y}_{t-1}$	0.4622	0.0666	0.4622	-0.4286	1.0000									
$\dot{Y}_{t-2}$	0.0744	0.1057	0.0744	0.0338	-0.4250	1.0000								
$\dot{Y}_{t-3}$	0.1110	0.0391	0.1110	-0.0779	0.0383	-0.4272	1.0000							
$\dot{Y}_{t-4}$	0.0282	0.0270	0.0282	-0.0013	-0.0861	0.0397	-0.4302	1.0000						
$\dot{Y}_{t-5}$	0.0360	-0.0280	0.0360	-0.0692	0.0048	-0.0990	0.0387	-0.4283	1.0000					
$\dot{Y}_{t-6}$	-0.0274	-0.0317	-0.0274	-0.0046	-0.0682	0.0096	-0.0955	0.0320	-0.4257	1.0000				
$\dot{Y}_{t-7}$	-0.0576	0.0794	-0.0576	0.1484	-0.0246	-0.0524	0.0015	-0.0789	0.0524	-0.4418	1.0000			
$\dot{Y}_{t-8}$	0.0874	0.0148	0.0874	0.0787	0.1544	-0.0322	-0.0516	-0.0002	-0.0888	0.0574	-0.4348	1.0000		
$\dot{Y}_{t-9}$	0.0191	0.0514	0.0191	0.0350	-0.0747	0.1560	-0.0284	-0.0585	0.0010	-0.0860	0.0452	-0.4326	1.0000	
$\dot{Y}_{t-10}$	0.0557	0.0765	0.557	0.0225	0.0390	-0.0712	0.1606	-0.0385	-0.0546	0.0036	-0.1058	0.0501	-0.4255	1.0000

Source: Box and Jenkins (1970, p. 525)

The second argument advanced for adding an additional number of lagged differences to the autoregressive equation was that the error terms e_t in the regression equation might be serially correlated. In order to determine if the recommendation based on this argument was valid, we obtained the residuals from the regression equation $Y_t = \alpha + \rho Y_{t-1} + e_t$ (Table 4 column 1) that did not include any lagged variable beyond Y_{t-1} and Figure 2 shows the graph of these residuals. Although not exactly normal distribution, these residuals follow an approximate normal distribution. Furthermore, we ran a Durbin-Watson test to determine if the residuals obtained from the regression equation $Y_t = \alpha + \rho Y_{t-1} + e_t$ had any serial correlation. For (2, 195) degrees of freedom (df), the Durbin-Watson d statistics was 1.967. We used the decision rule to fail to reject the null hypothesis of no serial correlation in the residual: $d_u < d^* < 4 -$

d_u. It corresponded to 1.56 < 1.967 < 2.44. The test results indicated that the null hypothesis could not be rejected and based on the available evidence no serial correlation in the residuals was detected. Clearly, these results were at odds with a justification to add additional number of lagged differences to the autoregressive equation on the ground of serial correlation in e_t.

Figure 2: Residual Analysis



We were not first to recognize the shortcomings in the study by Said and Dickey. Previously, Davidson and Mackinnon (2004, p. 620) and Hayashi (2000, p. 585) also identified these weaknesses. In addition, Perron (1989), Schwert (1989), Campbell and Perron (1991), Harris (1992), Taylor (2000), and Ng and Perron (2001) also noted these weaknesses. Interestingly, in the earlier development of tests for unit root, many writers held to the view that adding additional lagged differences to the autoregressive equation would be an appropriate solution as this would address both the serial correlation problem in e_t and the dependence that remained for most macroeconomic time series even after differencing. Based on new evidence, we found that adding additional lagged differences to the autoregressive equation on both counts would be inappropriate practice due to the threat to external validity. That is, the statistical inferences and conclusions based on the ADF test, albeit internally valid in the case being examined here, cannot be generalized to other populations and settings.

Conclusion

In this research note, we reanalyzed data reported by Box and Jenkins (1970) that were previously analyzed by Said and Dickey (1984, p. 606-607). Our data analysis reaffirms the internal validity of the research design in the study by Said and Dickey; however, we found their results to be incomplete. We reported new results from the reanalysis of original data not included in their study and showed the unit root finding to be sensitive to the length of the time lag included in the data analysis. Although the results obtained from the ADF test were valid in the study by Said and Dickey, those results could not be generalized to other population and settings. Put simply, the ADF test procedure leaves room open for the possibility that researchers can pick and choose the length of the time lag to obtain desired results. Thus, we cast doubt on the continued usefulness of the ADF test as a sound scientific method.

Acknowledgement

The research was presented at the 14th Annual International Conference on Global Studies & Athens International Symposium on Education and Research, December 19, 2020, at the International Confederation of Associations for Pluralism in Economics (ICAPE) Conference, January 3, 2019, Agnes Scott College, Atlanta, Georgia, USA, at the 83rd International Atlantic Economic Conference, March 22-25, 2017, Berlin, Germany, at the 86th Annual Meetings of the Southern Economic Association, November 19-21, 2016, Washington DC, USA, and at the 50th Annual Conference of the Canadian Economics Association, June 3 - 5, 2016, Ottawa, Canada. We thank Afshin Amiraslany, Krishna Kadiyala, Brandon Mackinnon, Mahadev Sharma and session participants in the above conferences for their helpful comments and suggestions. We also contacted Professor David Dickey of North Carolina State University, USA to request whether he liked to be acknowledged in this research. However, he declined our request.

Disclosure Statement

We declare that we did not apply any funding to conduct this research. We do not have any conflict of interest.

REFERENCES

Anderson, T. W. 1959. “On Asymptotic Distributions of Estimates of Parameters of Stochastic Difference Equations.” *Annals of Mathematical Statistics* 30(3): 676-687.

Banerjee, Anindya, Juan Dolado, John W. Galbraith, and David F. Hendry. 1993. *Co-Integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data.* Oxford: Oxford University Press.

Berk, Kenneth N. 1974. “Consistent autoregressive spectral estimates.” *The Annals of Statistics* 2(3): 489-502.

Bierens, Herman J. 2001. “Unit Roots.” Ch. 29 in *A Companion to Econometric Theory,* ed. Badi H. Baltagi, Blackwell Publishers, 610-633.

Box, George E. P. and Gwilym M. Jenkins. 1970. *Time Series Analysis Forecasting and Control.* San Francisco: Holden-Day.

Burman, Leonard E., W. Robert Reed, and James Alm. 2010. “A Call for Replication Studies.” *Public Finance Review* 38(6): 787-793.

Campbell, John Y., and Pierre Perron. 1991. “Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots.” *NBER Macroeconomics Annual* 6: 141 – 201.

Chang, Andrew C., and Phillip Li. 2015. “Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say “Usually Not”.” Finance and Economics Discussion Series 2015-083. Washington: Board of Governors of the Federal Reserve System.

Chinn, Menzie David. 1991. “Beware of econometricians bearing estimates: Policy analysis in a “unit root” world.” *Journal of Policy Analysis and Management* 10(4): 546 -567.

Clemens, Michael. 2017. “The Meaning of Failed Replications: A Review and Proposal.” *Journal of Economic Surveys* 31(1): 326–342.

Cochrane, John H. 1991. "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots: Comment" *NBER Macroeconomics Annual* 6: 201 – 210.

Davidson, James E. H., David B. Stephenson and Alemtsehai A. Turasie. 2016. "Time series modeling of paleoclimate data." *Environmetrics,* 27(1): 55–65.

Davidson, Russell and James G. Mackinnon. 2004. *Econometric Theory and Methods,* Oxford: Oxford University Press.

Dickey, David A., and Wayne A. Fuller. 1979. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association* 74(366): 427-431.

Dickey, David A., and Wayne A. Fuller. 1981. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." *Econometrica* 49(4): 1057-1072.

Diebold, F. and M. Nerlove. 1990. Unit Roots in in Economic Time Series: A Selective Survey. In T. Bewley (ed.), *Advances in Econometrics,* Vol. 8. Greenwich, CT: JAI Press.

Duvendack, Maren, Richard W. Palmer-Jones, and W. Robert Reed, 2015. "Replication in Economics. A Progress Report." *Econ Journal Watch* 12(2): 164–191.

Friedman, Milton. 1957. *A Theory of Consumption Function.* Princeton: Princeton University Press.

Fuller, W. A. 1976. *Introduction to Statistical Time Series.* New York: Wiley.

Fuller, W. A. 1984. "Nonstationary autoregressive time series." In Handbook of Statistics, 5, Time Series in the Time Domain, Ed. E. J. Hannan, P. R. Krishnaiah and M. M. Rao, Ch. 1, pp. 1-23. Amsterdam: North Holland.

Greene, William H. 2000. *Econometric Analysis,* 4th edition. New York: Macmillan.

Guilkey, David K., and Peter Schmidt. 1989. "Extended tabulations for Dickey-Fuller tests." *Economics Letters* 31(4): 355–357.

Hamilton, James D. 1994. *Time Series Analysis,* Princeton: Princeton University Press.

Hammermesh, Daniel S. 2007. Viewpoint: Replication in Economics. *Canadian Journal of Economics* 40: 715-733.

Harris, R. I. D. 1992. "Testing for Unit Roots Using the Augmented Dickey-Fuller Test: Some Issues Relating to the Size, Power and Lag Structure of the Test." *Economics Letters* 38 (4): 381-386.

Harvey, Andrew. 1997. "Trends, Cycles and Autoregressions." *The Economic Journal* 107(440): 192-201.

Hayashi, Fumio. 2000. *Econometrics,* Princeton: Princeton University Press.

Kaufmann, Robert K., Heikki Kauppi, and James H. Stock. 2006. "Emissions, concentrations, and temperature: a time series analysis." *Climatic Change* 77: 249-278.

Kaufmann, Robert K., Heikki Kauppi, Michael L. Mann, and James H. Stock. 2013. "Does temperature contain a stochastic trend: linking statistical results to physical mechanisms." *Climatic Change* 118: 729-743.

Libanio, Gilberto A. 2005. "Unit roots in macroeconomic time series: theory, implications, and evidence." *Nova Economia* 15(3): 145-176.

Luitel, Hari S. and Gerry J. Mahar. 2015a. "Why most published results on unit root and cointegration are false." *Ethical Economic Support Occasional Newsletter* of the Association for Integrity and Responsible Leadership in Economics and Associated Professions, December 22, 2015.

Luitel, Hari S. and Gerry J. Mahar. 2015b. "A short note on the application of Chow test of structural break in US GDP." *International Business Research,* 8(10), 112-116.

Luitel, Hari S. and Gerry J. Mahar. 2016. "Algebra of Integrated Time Series: Evidence from Unit Root Analysis." *International Advances in Economic Research,* 22(2), 199-209.

Luitel, Hari S., Gerry J. Mahar, Krishna Kadiyala, Daniel Friyia, and Brandon Mackinnon. 2018. "Are the Economies of Canada and the United States Integrated? Evidence from Cointegration Analysis." *American Review of Canadian Studies,* 48(1), 63-79.

Luitel, Hari and Gerry Mahar. 2021. "Why Economists Disagree: An Illustration of Irreconcilability Using the US State Level Unemployment Rate Data" in *Mimeo*, Algoma University, December 2021.

MacKinnon, James G. 1991. "Critical Values for Cointegration Tests." Chapter 13 in *Long-Run Economic Relationships: Readings in Cointegration*, ed. R. F. Engle and C. W. J. Granger, Oxford: Oxford University Press, 267-76.

MacKinnon, James G. 1994. "Approximate Asymptotic Distribution Functions for Unit-Root and Cointegration Tests." *Journal of Business & Economic Statistics* 12(2): 167-176.

Maddala, G. 1992. *Introduction to Econometrics,* 2nd edition. New York: Macmillan.

Maddala, G. S. and In-Moo Kim. 1998. *Unit Roots, Cointegration, and Structural Change.* Cambridge: Cambridge University Press.

Mann, H. B. and A. Wald. 1943. "On the Statistical Treatment of Linear Stochastic Difference Equations." *Econometrica* 11(3/4): 173-220.

McCullough, Bruce D. 2009. "Open Access Economics Journals and the Market for Reproducible Economic Research." *Economic Analysis and Policy,* 39(1): 117– 126.

McCullough, Bruce D., and Ross R. McKitrick. 2009. "Check the Numbers: The Case for Due Diligence in Policy Formation." *Studies in Risk & Regulation,* February, The Fraser Institute.

McCullough, Bruce D., Kerry Anne McGeary, and Teresa D. Harrison, 2008. "Do economics journal archives promote replicable research?" *Canadian Journal of Economics,* 41(4): 1406–1420.

Miguel, Eedward. 2015. "How Transparent is Economics Research?" Committee on the Status of Women in the Economics Profession (CSWEP) News, American Economic Association, 2015 Issue III.

Modigliani, Franco. 1986. "Life Cycle, Individual Thrift, and the Wealth of Nations." *The American Economic Review* 76(3): 297-313.

Moosa, Imad. 2011. "The Failure of Financial Econometrics: Assessing the Cointegration "Revolution"." *The Capco Institute Journal of Financial Transformation,* Applied Finance # 32.

Nelson, Charles R., and Charles R. Plosser. 1982. "Trends and Random Walks in Macroeconomic Time Series." *Journal of Monetary Economics* 10 (2): 139-162.

Ng, Serena and Pierre Perron. 2001. "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power." *Econometrica,* 69(6): 1519-1554

Patterson, Kerry. 2011. Unit Root Tests in Time Series Volume 1: Key Concepts and Problems, Palgrave Macmillan.

Patterson, Kerry. 2012. Unit Root Tests in Time Series Volume 2: Extensions and Developments, Palgrave Macmillan.

Perron, Pierre and Xiaokang Zhu. 2005. "Structural Breaks with Deterministic and Stochastic Trends." *Journal of Econometrics* 129: 65-119.

Perron, Pierre. 1989. "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis." *Econometrica* 57(6): 1361-1401.

Perron, Pierre. 1990. "Testing for a Unit Root in a Time Series Regression with a Changing Mean." *Journal of Business and Economic Statistics* 8(2): 153-162.

Phillips, Peter C. B. 1995. "Nonstationary Time Series and Cointegration." *Journal of Applied Econometrics* 10: 87-94.

Phillips, Peter C. B. 2003. "Laws and Limits of Econometrics." *The Economic Journal* 113(486): C26-C52.

Phillips, Peter C. B. and Zhijie Xiao. 1998. "A Primer on Unit Root Testing." *Journal of Economic Surveys* 12: 423–470.

Rao, M. M. 1961. "Consistency and Limit Distributions of Estimators of Parameters in Explosive Stochastic Difference Equations." *Annals of Mathematical Statistics* 32(1): 195-218.

Rothstein, Hannah R., Alexander J. Sutton and Michael Borenstein. 2005. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments.* John Wiley & Sons, Ltd.

Rubin, H. 1950. "Consistency of Maximum Likelihood Estimators in the Explosive Case." In T. Koopmans, ed. *Statistical Inference in Dynamic Economic Models.* New York: John Wiley and Sons.

Said, Said E., and David A. Dickey. 1984. "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order." *Biometrika* 71 (3): 599-607.

Schmidt, P. 1990. Dickey-Fuller Test with Drift. In T. B. Fomby and G. F. Rhodes (eds.), *Advances in Econometrics,* Vol. 8. Greenwich, CT: JAI Press, 161-200.

Schwert, G. William. 1989. "Test for Unit Roots: A Monte Carlo Investigation." *Journal of Business and Economic Statistics* 7 (2): 147-159.

Stock, James H. 1994. "Unit Roots, Structural Breaks, and Trends." Ch. 46 in R. Engle and D. McFadden (eds.), *Handbook of Econometrics,* volume IV. Amsterdam: Elsevier, 2740-2843

Stock, James H., and Mark W. Watson. 2003. *Introduction to Econometrics,* Pearson Education, Inc.

Storelvmo, T., T. Leirvik, U. Lohmann, P. C. B. Phillips and M. Wild. 2016. "Disentangling greenhouse warming and aerosol cooling to reveal Earth's climate sensitivity." *Nature Geoscience* 9(4): 286 – 289.

Taylor, A. M. Robert. 2000. "The Finite Sample Effects of Deterministic Variables on Conventional Methods of Lag-selection in Unit Root Tests." *Oxford Bulletin of Economics and Statistics* 62 (2): 293-304.

Titterington, D. M. 2013. "*Biometrika* highlights from volume 28 onwards." *Biometrika* 100 (1): 17-73.

White, John S. 1958. "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case." *Annals of Mathematical Statistics* 29(4): 1188-1197.

White, John S. 1959. "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case II." *Annals of Mathematical Statistics* 30(3): 831-834.

Zimmermann, Christian. 2015. "On the need for a replication journal." Federal Reserve Bank of St Louis, Working Paper 2015-016A.