NBER TECHNICAL WORKING PAPER SERIES

TESTS FOR UNIT ROOTS: A MONTE CARLO INVESTIGATION

G. William Schwert

Technical Working Paper No. 73

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 1988

This paper is an extension of an earlier working paper entitled "Effects of Model Specification on Tests for Unit Roots." Comments from Charles Nelson, Adrian Pagan, Peter Phillips, Charles Plosser, Peter Schmidt, Clifford Smith, Mark Watson, Ross Watts, anonymous referees, and especially from David Dickey and James Stock, were particularly helpful. Paul Seguin provided valuable computational assistance. The Bradley Policy Research Center at the University of Rochester provided support for this research. This research is part of NBER's research program in Financial Markets and Monetary Economics. Any opinions expressed are those of the author not those of the National Bureau of Economic Research.

NBER Technical Working Paper #73
December 1988

TESTS FOR UNIT ROOTS: A MONTE CARLO INVESTIGATION

ABSTRACT

Recent work by Said and Dickey(1984,1985), Phillips(1987), and Phillips and Perron(1988) examines tests for unit roots in the autoregressive part of mixed autoregressive-integrated-moving average (ARIMA) models (tests for stationarity). Monte Carlo experiments show that these unit root tests have different finite sample distributions than the unit root tests developed by Fuller(1976) and Dickey and Fuller(1979,1981) for autoregressive processes. In particular, the tests developed by Phillips(1987) and Phillips and Perron(1988) seem more sensitive to model misspecification than the high order autoregressive approximation suggested by Said and Dickey(1984).

G. William Schwert
William E. Simon Graduate School
of Business Administration
University of Rochester
Rochester, NY 14627

1. INTRODUCTION

Fuller(1976) and Dickey and Fuller(1979,1981) develop several tests of whether a p^{th} order autoregressive (AR) process,

$$Y_t = \alpha + \sum_{i=1}^{p} \phi_i Y_{t-i} + u_t, \tag{1}$$

is stationary. Stationarity implies that the roots of the lag polynomial $\phi(L) = (1 - \phi_1 L - \ldots \phi_p L^p)$ lie outside the unit circle (see Box and Jenkins(1976) for a discussion of stationarity in the context of AR processes). The null hypothesis in these tests is that the AR process contains one unit root, so the sum of the autoregressive coefficients in (1) equals 1. Dickey and Fuller estimate the model

$$Y_t = \alpha + \rho_\mu Y_{t-1} + \sum_{i=1}^{(p-1)} \phi_i' DY_{t-i} + u_t, \tag{2}$$

where $DY_{t-i} = Y_{t-i} - Y_{t-i-1}$, which is equivalent to the AR model in (1), except the coefficient ρ_μ should equal 1.0 if there is a unit root. Dickey and Fuller use Monte Carlo experiments to tabulate the sampling distribution of the regression "t-statistic," $\tau_\mu = (\hat{\rho}_\mu - 1)/s(\hat{\rho}_\mu)$, where $s(\hat{\rho}_\mu)$ is the standard error of the estimate $\hat{\rho}_\mu$ calculated by least squares. The distribution is skewed to the left and has too many large negative values relative to the Student-t distribution. See Dickey, Bell and Miller(1986) for a recent discussion of autoregressive unit root tests. Plosser and Schwert(1977) discuss a similar problem that arises when there is a unit root in the moving average polynomial. This can occur when differencing is used to remove nonstationarity and the true model is a stationary and invertible ARMA model around a time trend.

This paper analyzes the sensitivity of the Dickey-Fuller tests to the assumption that the time series is generated by a pure autoregressive process.

In particular, when a variable is generated by a mixed ARIMA process, the critical values implied by the Dickey-Fuller simulations can be misleading. Section 2 describes recent extensions of the Dickey-Fuller test procedure suggested by Said and Dickey(1984, 1985), Phillips(1987), Phillips and Perron(1988), and Perron(1986a,b) that attempt to account for mixed ARIMA processes as well as pure AR processes in performing unit root tests. Section 3 contains results of a Monte Carlo experiment that calculates the size of the Dickey-Fuller and the related test statistics when the true process is ARIMA rather than AR. Section 4 contains concluding remarks.

2. EXTENSIONS OF THE DICKEY-FULLER TESTS

Said and Dickey(1984) argue that an unknown ARIMA(p,1,q) process can be adequately approximated by an ARIMA(k,0,0) process, where $k=o(T^{1/3})$. Given this approximation, the limiting distribution of the unit root test based on a high order AR approximation will be the same as the Dickey-Fuller distribution. Of course, for a given application this argument does not indicate the appropriate number of lags k.

To understand why a finite order autoregressive process may not provide an adequate approximation to a mixed ARIMA process, it is useful to consider the infinite order autoregressive process implied by an ARIMA(0,1,1) process for different values of the moving average parameter θ. The autoregressive coefficients are calculated by matching coefficients of the lag operator L in the relations

$$\pi(L) = (1 - L)/(1 - \theta L) \Rightarrow (1 - \theta L) \quad \pi(L) = (1 - L),$$

where π_i is the autoregressive coefficient at lag i. The autoregressive coefficients decay slowly for large absolute values of the moving average parameter. The sum of the coefficients is equal to unity (the value for the infinite sum of all autoregressive coefficients for this nonstationary process) to four decimal places after 24 lags for θ equal to .5 or -.5. For values of θ equal to .8, .9 and .95, however, the sums of the coefficients to 24 lags are equal to .9953, .9202, and .7080, respectively. This suggests that the approximation error caused by estimating a finite order AR process is large for moving average parameters greater than .8. Such series have autocorrelations for the levels of the series that decay slowly, and first order autocorrelations for the first differences close to -.50 (see Wichern(1973) and Schwert(1987)).

Said and Dickey(1985) show that the unit root estimator from an ARIMA(1,0,1) process,

$$Y_t = \alpha + \rho_\mu Y_{t-1} + u_t - \theta u_{t-1}, \tag{3}$$

has the asymptotic distribution tabulated by Dickey and Fuller when one Gauss-Newton step is taken from initial values $\rho_\mu=1$ and θ equal to a consistent estimator conditional on $\rho_\mu=1$. They provide limited Monte Carlo evidence that shows the effect of estimating the moving average parameter θ on the unit root test statistic τ_μ.

Fuller(1976, p. 371) presents fractiles of the distribution of $T(\hat{\rho}_\mu-1)$ when $\rho_\mu=1$ and $\alpha=0$ for an ARIMA(1,0,0) process,

$$Y_t = \alpha + \rho_\mu Y_{t-1} + u_t, \qquad t=1, \ldots T. \tag{4}$$

This normalized measure of bias provides another test of the unit root hypothesis. Dickey and Fuller(1979) show that tests based on this statistic are more powerful against the alternative hypothesis that $\rho_\mu<1$ than the test based on the τ_μ statistic.

The distribution of the estimator $\hat{\rho}_\mu$ depends on the structure of the ARIMA process that generated the data. As noted by Fuller(1976, pp. 373-382), the statistic $Tc(\hat{\rho}_\mu-1)$ from a general ARIMA model has the same distribution as $T(\hat{\rho}_\mu-1)$ from the AR(1) model, where the constant c is the sum of the coefficients ψ_i from the moving average representation of the errors from (4). $\psi(L)=\theta(L)/\phi(L)$. One strategy for estimating the constant c is to use the additional coefficients from the ARIMA(p,0,0) model in (2), or from an ARIMA(p,0,q) model, where ϕ_i' are the (p-1) autoregressive coefficients for DY_{t-i}.

Phillips(1987) and Phillips and Perron(1988) also show that the Dickey-Fuller tests are affected by autocorrelation in the errors from (4). They develop modifications of the test statistics τ_μ and $T(\hat{\rho}_\mu-1)$ that have the asymptotic distributions tabulated by Dickey and Fuller, when the data follow an ARIMA(p,0,q) process. In fact, these papers allow for more general dependence in the error process, including conditional heteroskedasticity. These adjustments involve the autocovariances of the errors from an ARIMA(1,0,0) model in (4). They modify the test statistic $T(\hat{\rho}_\mu-1)$,

$$Z_{\rho\mu} = T(\hat{\rho}_\mu-1) - .5(s_{T\ell}^2-s_u^2)\ T^2\ \{\sum_{t=2}^{T}(Y_{t-1} - \bar{Y}_{-1})^2\}^{-1}, \tag{5}$$

where s_u^2 is the sample variance of the residuals u_t,

$$s_{T\ell}^2 = T^{-1}\sum_{t=1}^{T}u_t^2 + 2T^{-1}\sum_{j=1}^{\ell}\omega_{j\ell}\sum_{t=j+1}^{T}u_t u_{t-j}\ , \tag{6}$$

and the weights $\omega_{j\ell} = \{1 - j/(\ell+1)\}$ ensure that the estimate of the variance $s_{T\ell}^2$ is positive (see Newey and West(1987)). Following the intuition of Said and Dickey(1984), they suggest that the number of lags ℓ of the residual autocovariances in (6) be allowed to grow with the sample size T.

Phillips and Perron modify the regression "t-test" τ_μ,

$$Z_{\tau\mu} = \tau_{\mu}(s_u/s_T) - .5(s_{T\ell}^2 - s_u^2)\ T\ \{s_{T\ell}^2 \sum_{t=2}^{T} (Y_{t-1} - \bar{Y}_{-1})^2\}^{-1/2} \tag{7}$$

where $s_{T\ell}^2$ is defined in (6).

Dickey and Fuller also consider tests with a time trend included as an additional regressor, so the alternative hypothesis is a stationary process around a time trend. Thus, the ARIMA(1,0,0) model in (4) is modified so

$$Y_t = \alpha + \beta\ [t-(T+1)/2] + \rho_\tau\ Y_{t-1} + u_t, \tag{8}$$

the ARIMA(1,0,1) model in (3) is modified so

$$Y_t = \alpha + \beta\ [t-(T+1)/2] + \rho_\tau\ Y_{t-1} + u_t - \theta\ u_{t-1}, \tag{9}$$

and the ARIMA(p,0,0) process in (2) is modified so

$$Y_t = \alpha + \beta\ [t-(T+1)/2] + \rho_\tau\ Y_{t-1} + \sum_{i=1}^{(p-1)} \phi_i'\ DY_{t-i} + u_t. \tag{10}$$

The regression "t-tests," τ_τ, are important because Evans and Savin(1984) show that τ_μ statistics are a function of the unknown intercept α in (2) or (4). On the other hand, including a time trend in (8), (9), or (10), even when the trend coefficient $\beta=0$, makes the distribution of the autoregressive parameter estimate $\hat{\rho}_\tau$ independent of α. In empirical applications, where knowledge of the value of the intercept α is unavailable, inclusion of a time trend is probably a prudent decision in performing unit root tests.

Phillips and Perron(1988) develop adjustments to the Dickey-Fuller tests $T(\hat{\rho}_\tau-1)$ and τ_τ where the alternative hypothesis is a stationary ARIMA(p,0,q) process around a deterministic time trend. They show that the test statistic,

$$Z_{\rho\tau} = T(\hat{\rho}_\tau-1) - (s_{T\ell}^2 - s_u^2)\ (T^6/24D_{XX}), \tag{11}$$

has the asymptotic distribution tabulated by Dickey and Fuller for $T(\hat{\rho}_\tau-1)$ in the ARIMA(1,0,0) case, where D_{XX} is the determinant of the regressor cross-product matrix. Their modification to the statistic r_τ is:

$$Z_{r\tau} = r_\tau(s_u/s_{T\ell}) - (s_{T\ell}^2 - s_u^2)\, T^3 \left\{ s_{T\ell}\; 4(3D_{XX})^{1/2} \right\}^{-1}. \qquad (12)$$

This statistic should have the asymptotic distribution tabulated by Dickey and Fuller for r_τ, even when the regression errors in (8) are autocorrelated.

3. A MONTE CARLO EXPERIMENT FOR UNIT ROOT TESTS

The Monte Carlo experiment examines the effects of model misspecification on the size of unit root tests for mixed ARIMA processes. The experiment constructs the data to follow an ARIMA(0,1,1) process,

$$Y_t = Y_{t-1} + u_t - \theta\, u_{t-1}, \qquad t=-19, \ldots T,$$

where the errors $\{u_t\}$ are serially uncorrelated standard normal variables. The data are generated by setting u_{-20} and Y_{-20} equal to 0 and creating T+20 observations, discarding the first 20 observations to remove the effect of the initial conditions. Samples of size T = 25, 50, 100, 250, 500, and 1000 are used in the experiments. Each experiment is replicated 10,000 times to create the sampling distribution for the test statistics. The moving average parameter θ is set equal to .8, .5, 0, -.5, and -.8, which implies first order autocorrelations for the first differences of these series of -.49, -.40, 0, .40 and .49. The first order autocorrelation coefficient for an ARIMA(0,0,1) process equals $-\theta/(1+\theta^2)$. Higher order autocorrelations equal 0.

3.1 Regression "t-tests"

Several tests of nonstationarity are performed on each data series. First, the regression "t-test" from (4) studied by Dickey and Fuller is calculated to illustrate the problems that occur when the data are generated by a process other than AR(1). Second, two versions of the Phillips and Perron(1988) test are calculated, using different numbers of lags ℓ of the residual autocorrelations in calculating $s^2_{T\ell}$ in (6):

$$\ell_4 = \text{Int}(4(T/100)^{1/4}) \qquad (13a)$$

$$\ell_{12} = \text{Int}(12(T/100)^{1/4}), \qquad (13b)$$

so ℓ_4=4 and ℓ_{12}=12 when T=100, (when T=25, ℓ_4=2 and ℓ_{12}=8; when T=1000, ℓ_4=7 and ℓ_{12}=21). Third, an ARIMA(1,0,1) model is estimated to test whether the autoregressive coefficient ρ_μ equals 1.0, using the "t-test," $\tau_\mu = (\hat{\rho}_\mu - 1)/s(\hat{\rho}_\mu)$, where $s(\hat{\rho}_\mu)$ is the standard error calculated by an iterative nonlinear least squares algorithm. Note that this is not the procedure suggested by Said and Dickey(1985), since their results require only one Gauss-Newton step from the unit root. Nevertheless, empirical researchers who estimate ARIMA(1,0,1) models and discover an estimated autoregressive parameter close to unity would want to know the reliability of the "t-test" for the unit root when iterative least squares is used. Fourth, an AR(ℓ_4) model is estimated in equation (2) and the regression "t-test" is used to test whether ρ_μ equals 1. Finally, an AR(ℓ_{12}) model is estimated in equation (2) to calculate τ_μ. The latter tests follow the suggestion of Said and Dickey(1984) to use a high order autoregressive process to approximate an unknown ARIMA process, where the order of the autoregression grows with the sample size T as in (13a,b).

Table 1a contains estimates of the sizes of tests using the 1% critical values from the Dickey-Fuller distribution for τ_μ, for the six different test statistics (AR(1); Phillips-Perron with ℓ_4 lags, $Z_{\tau\mu}(\ell_4)$; Phillips-Perron with ℓ_{12} lags, $Z_{\tau\mu}(\ell_{12})$; ARIMA(1,0,1); AR(ℓ_4); and AR(ℓ_{12})), for the six different sample sizes (T = 25, 50, 100, 250, 500, and 1000), and for the five different values of the moving average parameter for the true process (θ = .8, .5, 0, -.5, and -.8), where the alternative hypothesis is a stationary ARMA process around a constant mean. Table 1b contains the estimates of the sizes of tests using the 5% critical values. These tables do not report the upper tail of the sampling distributions because the usual alternative hypothesis is that the process is stationary ($\rho_\mu < 1$). As previously reported by Dickey and Fuller, the distribution of the τ_μ statistics has a negative mean and is skewed toward negative values for all of the cases considered in these experiments. Additional information about these sampling distributions is available from the author on request. The simulations were programmed in FORTRAN using the IMSL subroutine GGNOF to generate pseudo-random normal variates. All results were checked using the RATS computer program.

The first thing to note about Tables 1a and 1b is that the simple AR(1) test is severely affected by the presence of moving average components in the data generation process. The estimated size for this test is positively related to the moving average parameter θ, being too large for θ=.5 or .8, and too small for θ=-.5 or -.8. Of course, this problem is exactly what motivates the tests proposed by Said and Dickey and by Phillips and Perron.

Second, the Phillips and Perron tests do not have distributions that are close to the Dickey-Fuller distribution, especially for θ=.5 or .8. At both the 1% and 5% levels, the size of the Phillips-Perron tests are much larger than the nominal size of the test, even for samples as large as T=1000. As the number of lags of the residual autocorrelations used in (6) increases from ℓ_4 to ℓ_{12} the size estimates become farther away from the Dickey-Fuller results. The Phillips-Perron tests are much closer to the Dickey-Fuller distribution for negative moving average parameters θ=-.5 and -.8, although the size is too small for these cases.

The second thing to note about Tables 1a and 1b is that estimating a moving average parameter along with the unit root changes the behavior of the sampling distribution for the test statistic. This is interesting because Dickey and Fuller show that asymptotically the unit root test τ_μ is not affected by estimation of higher order autoregressive parameters. Said and Dickey(1985) show that the asymptotic behavior of the unit root test should not be affected by the estimation of moving average parameters when only one iterative step is taken from the unit root. For positive values of the moving average parameter θ, the size of the ARIMA(1,0,1) test is above the nominal size based on the Dickey-Fuller distribution. This difference is largest for both small (T=25 or 50) and large (T=500 or 1000) sample sizes, with the size being closest for moderate sample sizes (T=100 or 250). The apparent lack of convergence to the Dickey-Fuller statistic as the sample size grows contrasts with the results of Said and Dickey(1985) who examine samples of 49 and 99 observations. Apparently, the distinction between the one-step method proposed by Said and Dickey versus the iterative estimation used in these experiments is important.

The tests based on the ℓ_4-order autoregressive model are close to the Dickey-Fuller results for values of the moving average parameter θ equal to .5, 0, -.5, or -.8. With θ equal to .8, however, the AR(ℓ_4) approximation is deficient in that the size of the test is well above the nominal size using the Dickey-Fuller distribution, although this problem seems to be reduced as the sample size grows.

The size estimates based on the ℓ_{12}-order autoregressive model are closer to the nominal size than for the AR(ℓ_4) model. The only notable departure from the Dickey-Fuller results is for θ equal to .8. In this case, with small sample sizes (T=25) the size of the AR(ℓ_{12}) test is below the nominal size based on the Dickey-Fuller distribution.

Tables 2a and 2b contain estimates of the size of unit root tests at the 1% and 5% levels, respectively, where the alternative hypothesis is a stationary ARMA process around a time trend. As noted by Dickey and Fuller, including a time trend causes the critical values of τ_τ to be lower than τ_μ (i.e., the regression t-statistic must be more negative to reject the unit root hypothesis). Nevertheless, the relative patterns in Tables 1a and 1b are repeated in Tables 2a and 2b. For example, the sizes of the ARIMA(1,0,1) test and of the AR(ℓ_4) test are above the nominal size based on the Dickey-Fuller critical values for θ=.8. As in Tables 1a and 1b, the higher order autoregressive approximation AR(ℓ_{12}) has size close to the nominal level for sample sizes greater than 50. The Phillips-Perron tests have sizes that are furthest from the nominal size, with the largest departures for cases where θ is positive. In fact, with θ=.8, the Phillips-Perron tests reject a unit root over 99% of the time for a nominal 1% level test for sample sizes greater than 50.

Thus, a low order autoregressive approximation can lead to misspecification of unit root tests when the moving average parameter is large. Higher order AR processes seem to mitigate the problem (although the order of the AR process necessary to provide an adequate approximation can be quite large for θ=.8 or higher). Unit root tests based on the mixed ARIMA(1,0,1) model require moderate sample sizes before the Dickey-Fuller fractiles are accurate.

3.2 The Distribution of the Normalized Unit Root Estimator

Tables 3a and 3b contain estimates of the size of tests based on the normalized unit root estimator $T(\hat{\rho}_\mu-1)$ at the 1% and 5% levels, respectively. Six different tests are considered (AR(1); Phillips-Perron with ℓ_4 lags, $Z_{\rho\mu}(\ell_4)$; Phillips-Perron ℓ_{12} lags, $Z_{\rho\mu}(\ell_{12})$; ARIMA(1,0,1); AR(ℓ_4) corrected using the estimated value of the autoregressive parameters; and AR(ℓ_{12}) corrected using the estimated value of the autoregressive parameters), for the six different sample sizes (T = 25, 50, 100, 250, 500, and 1000), and for the five different values of the moving average parameter (θ = .8, .5, 0, -.5, and -.8), where the alternative hypothesis is a stationary ARMA process around a constant mean.

In many ways the results in Tables 3a and 3b are easier to summarize than the results in Tables 1a, 1b, 2a and 2b. For the AR(1) model, the estimated size is above the nominal level for θ equal to .8 and .5, and the difference increases with the sample size. The corrections suggested by Phillips and Perron(1988) do not reduce this problem much, and the use of more lags ℓ_{12} harms the performance of the test in this case.

The results for the ARIMA(1,0,1) model are interesting. For negative values of θ, the size is close to the nominal size from the Dickey-Fuller distribution for all sample sizes. For positive values of θ, the estimated size is higher than the nominal size for all sample sizes. Unfortunately, I did not compute the 'corrected' version of this test, $T(1-\theta)(\hat{\rho}_\mu-1)$, but such a correction would probably have improved the performance of this test substantially.

The AR(ℓ_4) test yields estimates of the size that are systematically related to the moving average parameter, θ. Higher values of θ yield lower estimates of the unit root, so the AR(ℓ_4) size estimates are well above the nominal size based on the Dickey-Fuller distribution when θ equals .8. The AR(ℓ_4) size estimates are too low when θ equals -.5 or -.8. These problems are reduced for larger sample sizes.

The AR(ℓ_{12}) test is better than the AR(ℓ_4) test for larger sample sizes, but worse for smaller sample sizes. For small sample sizes (25 and 50), the larger number of parameters that must be estimated in the AR(ℓ_{12}) model apparently bias the unit root estimator downward. Note that even when the moving average parameter θ equals zero, so the true process is a random walk as originally assumed by Dickey and Fuller, the estimated size for the AR(ℓ_{12}) test is well above the nominal size of the test. For large samples (T=250 or above), the sizes are closer to the nominal level of the tests, although they are still too high.

Tables 4a and 4b contain estimates of the size of tests based on the normalized unit root estimator $T(\hat{\rho}_\tau-1)$ at the 1% and 5% levels, respectively, where the alternative hypothesis is a stationary ARMA process around a time trend. The relative patterns in Tables 4a and 4b are virtually identical to those in Tables 3a and 3b. As noted by Fuller(1976), the size of the Dickey-Fuller tests is related to the moving average parameter θ. When θ=.8, the estimated size is far above the nominal level of the test. The corrections suggested by Fuller stabilize the behavior of the statistic for different values of θ, although the size of these tests is above the nominal size using the Dickey-Fuller distribution. The corrections suggested by Phillips and Perron(1988) do not work as well, since the estimated size remains well above the nominal size for positive values of θ.

The effects of model misspecification are clearer in the normalized bias tests (Tables 3a, 3b, 4a and 4b) than in the "t-tests" (Tables 1a, 1b, 2a and 2b). When the data are generated by an integrated moving average process, high order autoregressive approximations yield biased estimates of the unit root coefficient. With positive moving average parameters the unit root coefficients are too small, and with negative moving average parameters the unit root coefficients are too large. Even though the results of Dickey-Fuller(1979) suggest that $T(\hat{\rho}_\mu-1)$ provides a more powerful test than the τ_μ statistic when $\rho_\mu<1$, the results above suggest than the τ_μ and τ_τ statistics are less sensitive to model misspecification. The corrections to the normalized unit root estimator suggested by Phillips(1987) and Phillips and Perron(1988) do not work well in the cases examined here. The corrections suggested by Fuller(1976) improve the behavior of the normalized unit root test for high order autoregressive models with very large sample sizes, but they distort the size of the test in small to moderate samples.

3.3 Further Analysis of the Phillips and Perron Tests

The Phillips(1987) and Phillips and Perron(1988) tests perform poorly in cases where the true data are generated by an ARIMA(0,1,1) processes with θ=.5 or θ=.8. This has been documented earlier by Monte Carlo experiments in Perron's dissertation(1986a), although the extent of the problem was not as clear in his work. Phillips and Perron(1988), in Monte Carlo work that postdated this paper, find results that are similar to the results above. It is surprising with sample sizes as large as 500 or 1000 that these tests are not close to the Dickey-Fuller distribution, as they should be in 'large samples.'

To provide further insight into this problem, additional Monte Carlo experiments are performed to analyze the Phillips-Perron tests, $Z_{\rho\mu}(l)$ and

$Z_{\tau\mu}(\ell)$. The procedure discussed above is used, except only the case with $\theta=.8$ is considered. Sample sizes of $T=1{,}000$ and $T=10{,}000$ are used. The number of residual autocorrelations ℓ used to calculate the variance $s^2_{T\ell}$ in (6) is varied from 0 (no adjustment) to ℓ_{12} ($\ell_4=7$ and $\ell_{12}=21$ when $T=1{,}000$; $\ell_4=12$ and $\ell_{12}=37$ when $T=10{,}000$). Table 5a contains the 5% and 1% fractiles of the sampling distributions from 10,000 replications for the Phillips-Perron test, $Z_{\rho\mu}(\ell)$. Table 5b contains the 5% and 1% fractiles of the sampling distributions from 10,000 replications for the Phillips-Perron test, $Z_{\tau\mu}(\ell)$. Tables 5a and 5b also contain the estimated size of the 5% and 1% level tests in parentheses below the estimated critical values.

There are two questions about the best way to do the Phillips-Perron tests. First, there is a question of the number of lags of the residual autocorrelations ℓ to use. Second, there is a question about the way to estimate the variances s^2_u and $s^2_{T\ell}$.

If the unit root estimate is equal to its true value, $\rho_\mu=1$, the residual autocorrelations should equal -0.49 at lag 1 and 0.0 at the remaining lags. For the data generating process used in these simulations, a relatively low number of lags should work best. Thus, Tables 5a and 5b show values of the Phillips-Perron tests based on $\ell=0$, 1, 2, 3, 4, ℓ_4 and ℓ_{12}, where $\ell=0$ is the original Dickey-Fuller statistic.

Phillips and Perron suggest two strategies for estimating the variances s^2_u and $s^2_{T\ell}$. The technique used in the simulations above is based on residuals from the estimate of (4), which is the procedure recommended in the first draft of the Phillips-Perron paper. The alternative procedure is to assume the autoregressive parameter ρ_μ equals one, and use the differences DY_t to calculate the variance estimates (a procedure also discussed by Phillips and Perron). This distinction is important because the autocorrelations of the residuals are not similar to the autocorrelations of the differences when $\theta=.8$. Because the estimate of the unit root $\hat{\rho}_\mu$ is well below one in most cases when $\theta=.8$, the residual autocorrelation at lag 1 averages -.367 when T=1,000, and the remaining autocorrelations are positive and decay very slowly (from .071 at lag 2 to .060 at lag 21). This is typical of a mixed ARIMA(1,0,2) process with an autoregressive coefficient close to unity. For an ARIMA(1,0,2) model, the k^{th} autocorrelation $\rho_k = \rho_2\phi^{k-2}$, where ρ_k is the autocorrelation at lag k and ϕ is the autoregressive parameter. Based on the estimates $r_2 = .071$ and $r_{21} = .060$, the implied value of ϕ is .99. These positive residual autocorrelations cause the Phillips-Perron tests to grow farther from the Dickey-Fuller distribution as more lags are included. Thus, the two-step procedure recommended by Phillips and Perron seems to have an important flaw: the estimate of the autoregressive root $\hat{\rho}_\mu$ in (4) is biased substantially below one when $\theta=.8$, so the residuals from (4) retain much of the nonstationarity from the original series.

In contrast, the average autocorrelation of the differences equals -.486 at lag 1, and equals .000 at all remaining lags when T=1,000. Nevertheless, the performance of the Phillips-Perron tests based on differences in Tables 5a and 5b seems to improve as the number of lags increases. This is probably due to the Newey-West weighting scheme used to calculate the variance estimate $s^2_{T\ell}$ in (6), which gives greater weight to the autocorrelation at lag 1 as the number of lags increases.

The results for samples of 10,000 observations in Tables 5a and 5b are closer to the Dickey-Fuller distribution than the results for samples of 1,000 observations, but the rate of convergence seems very slow. Finally, with samples of T=10,000, using residuals to calculate the variance estimates, the Phillips-Perron test based on ℓ_4=12 lags exhibits unusual behavior. For example, the .05 critical values for $Z_{\rho\mu}(\ell)$ is above the Dickey-Fuller critical value, although the .01 critical value is below the Dickey-Fuller value.

Based on the results in Tables 5a and 5b, the size of the Phillips-Perron tests is better specified when using differences to calculate the variance estimates if θ=.8, although the Said-Dickey tests are closer to the Dickey-Fuller distribution. One should be cautious, however, before concluding that one should always use differences in the Phillips-Perron test. In discussing the multivariate analog to the Phillips-Perron test $Z_{\rho\mu}(\ell)$, Stock and Watson(1987) show that this test is not consistent versus some stationary alternative hypotheses when using the differences to calculate the variance estimates. Thus, the Phillips-Perron tests using residuals behave poorly under the null hypothesis, but the tests based on the differences behave poorly under some plausible alternative hypotheses.

4. SUMMARY

The ARIMA(1,0,1) process used in the Monte Carlo experiments approaches a stationary random process as the moving average parameter θ approaches the autoregressive parameter ρ_μ. For cases where ρ_μ is close to or equal to one, and θ is less than but close to ρ_μ, the autocorrelations of the data are small positive numbers that decay very slowly. These cases occur frequently in economic data. For example, Nelson and Schwert(1977) find the monthly C.P.I. inflation rate for the U.S. follows such a process; Huberman and Schwert(1985) find that the monthly Israeli C.P.I. inflation rate follows such a process; French, Schwert and Stambaugh(1987) find that the log of monthly stock market volatility follows such a process. Schwert(1987) applies the unit root tests discussed in this paper to 17 important U.S. macroeconomic time series and concludes that many of the tests would falsely reject the unit root hypothesis using the Dickey-Fuller critical values. In such cases, the common argument that the unit root in the autoregressive part of the model dominates the asymptotic behavior of the process is misleading for large finite samples.

The simulations in this paper show that the tests for unit roots developed by Dickey and Fuller are sensitive to the assumption that the data are generated by a pure autoregressive process. When the underlying process contains a moving average component, the distribution of the unit root test statistics can be far different from the distributions reported by Dickey and Fuller. Moreover, the tests recently suggested by Said and Dickey(1984,1985), Phillips(1987) and Phillips and Perron(1988) to correct the model misspecification problem do not seem to work well when the moving average parameter is large. In particular, the tests proposed by Phillips and Perron do not come close to their asymptotic distribution for samples as large as 10,000 observations. The best test, in the sense that it has size close to its nominal level for all values of the moving average parameter θ, is the Said and Dickey(1984) high order autoregressive "t-test" for the unit root.

Given the many reasons to believe that economic time series contain moving average components, these simulation experiments provide warning against the broad application of unit root tests in economics. It is important to consider the correct specification of the ARIMA process before testing for the presence of a unit root in the autoregressive polynomial.

REFERENCES

Box, G. E. P. and Jenkins, G. M. (1976), Time Series Analysis: Forecasting and Control(Rev. ed.), San Francisco, CA: Holden-Day.

Dickey, D. A., Bell, W. R. and Miller, R. B. (1986), "Unit Roots in Time Series Models: Tests and Implications," The American Statistician, 40, 12-26.

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

Dickey, D. A., and Fuller, W. A. (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," Econometrica, 49, 1057-1072.

Evans, G. B. A., and Savin, N. E. (1984), "Testing for Unit Roots: 2," Econometrica, 52, 1241-1269.

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987), "Expected Stock Returns and Volatility," Journal of Financial Economics, 19, 3-29.

Fuller, W. A. (1976), Introduction to Statistical Time Series. New York: John Wiley.

Huberman, G., and Schwert, G. W. (1985), "Information Aggregation, Inflation and the Pricing of Indexed Bonds," Journal of Political Economy, 93, 92-114.

Nelson, C. R., and Plosser, C. I. (1982), "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," Journal of Monetary Economics, 10, 139-162.

Nelson, C. R., and Schwert, G. W. (1977), "On Testing the Hypothesis That the Real Rate of Interest Is Constant," American Economic Review, 67, 478-486.

Newey, W. K., and West, K. D. (1987), "A Simple Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," Econometrica, 55, 703-708.

Perron, P. (1986a), "Hypothesis Testing in Time Series Regression with a Unit Root," unpublished Ph.D. dissertation, Yale University, Dept. of Economics.

Perron, P. (1986b), "Trends and Random Walks in Macroeconomic Time Series: Further Evidence from a New Approach," manuscript, University of Montreal, Dept. of Economics.

Phillips, P. C. B. (1987), "Time Series Regression with a Unit Root," Econometrica, 55, 277-301.

Phillips, P. C. B., and Perron, P. (1988), "Testing for a Unit Root in Time Series Regression," Biometrika(forthcoming).

Plosser, C. I., and Schwert, G. W. (1977), "Estimation of a Noninvertible Moving Average Process: The Case of Overdifferencing," Journal of Econometrics, 6, 199-224.

Poterba, J. M., and Summers, L. H. (1986), "The Persistence of Volatility and Stock Market Fluctuations," American Economic Review, 76, 1142-1151.

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

Said, S. E., and Dickey, D. A. (1985), "Hypothesis Testing in ARIMA(p,1,q) Models," Journal of the American Statistical Association, 80, 369-374.

Schwert, G. W. (1987), "Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data," Journal of Monetary Economics, 20, 73-103.

Stock, J. and Watson, M. (1987), "Testing for Common Trends," Hoover Institution Discussion Paper No. E-87-2.

Wichern, D. W. (1973), "The Behavior of the Sample Autocorrelation Function for an Integrated Moving Average Process," Biometrika, 60, 235-239.

Table 1a. Empirical Size for 1% Level Test

Based on Dickey-Fuller Distribution of τ_μ for ρ_μ=1

Sample Size, T (DF crit. value)	Moving Avg Parameter θ	AR(1)	$Z_{\tau\mu}(\ell_4)$	$Z_{\tau\mu}(\ell_{12})$	ARMA(1,1)	AR(ℓ_4)	AR(ℓ_{12})
25 (-3.75)	.8	.722	.719	.745	.061	.227	.007
	.5	.196	.193	.213	.053	.040	.008
	.0	.009	.010	.015	.022	.008	.008
	-.5	.007	.006	.014	.025	.023	.010
	-.8	.007	.004	.012	.044	.031	.012
50 (-3.58)	.8	.952	.938	.975	.068	.220	.009
	.5	.312	.277	.376	.024	.020	.007
	.0	.010	.011	.011	.005	.009	.007
	-.5	.007	.006	.008	.010	.008	.008
	-.8	.006	.004	.006	.032	.005	.009
100 (-3.51)	.8	.982	.962	.988	.037	.216	.011
	.5	.374	.291	.417	.005	.014	.008
	.0	.011	.012	.012	.008	.011	.010
	-.5	.005	.005	.004	.010	.012	.010
	-.8	.008	.007	.007	.019	.021	.009
250 (-3.46)	.8	.992	.952	.981	.021	.194	.010
	.5	.422	.247	.366	.048	.014	.009
	.0	.011	.011	.012	.029	.009	.008
	-.5	.005	.005	.005	.020	.009	.009
	-.8	.008	.007	.006	.023	.009	.009
500 (-3.44)	.8	.993	.925	.968	.107	.231	.012
	.5	.437	.185	.286	.050	.014	.009
	.0	.011	.012	.012	.028	.010	.009
	-.5	.005	.007	.007	.019	.009	.010
	-.8	.006	.007	.006	.020	.007	.009
1000 (-3.43)	.8	.994	.887	.941	.134	.100	.010
	.5	.442	.139	.218	.055	.011	.010
	.0	.009	.009	.010	.024	.009	.009
	-.5	.005	.008	.007	.021	.010	.009
	-.8	.006	.009	.007	.023	.009	.010

NOTE: the proportion of statistics less than the 1% critical value from Fuller[1976, p. 373, Table 8.5.2] for the regression "t-test" for a unit root τ_μ against the alternative hypothesis that the process is stationary around a constant mean. Based on 10,000 replications of an ARIMA(0,1,1) process,

$$(Y_t - Y_{t-1}) = \epsilon_t - \theta\, \epsilon_{t-1},\quad t=1,\ldots T.$$

The Dickey-Fuller critical values are in parentheses under the sample size. The AR(1) test is based on equation (4); the Phillips corrections to the AR(1) test, $Z_{\tau\mu}$, use equations (6) and (7); the ARMA(1,1) test uses equation (3); the AR(ℓ_4) and AR(ℓ_{12}) tests use equation (2) with ℓ_4 and ℓ_{12} lags, respectively, where ℓ_4 and ℓ_{12} are defined in (13a,b). The standard error for these estimates of the size of the tests is .001.

Table 1b. Empirical Size for 5% Level Test

Based on Dickey-Fuller Distribution of τ_μ for $\rho_\mu=1$

Sample Size, T (DF crit. value)	Moving Avg Parameter θ	AR(1)	$Z_{\tau\mu}(\ell_4)$	$Z_{\tau\mu}(\ell_{12})$	ARMA(1,1)	AR(ℓ_4)	AR(ℓ_{12})
	.8	.923	.919	.925	.094	.522	.036
	.5	.418	.400	.436	.076	.143	.038
25	.0	.050	.051	.055	.037	.052	.039
(-3.00)	-.5	.030	.028	.039	.049	.090	.046
	-.8	.029	.024	.035	.085	.111	.051
	.8	.989	.980	.994	.098	.471	.046
	.5	.523	.454	.557	.038	.082	.035
50	.0	.051	.053	.049	.020	.047	.036
(-2.93)	-.5	.027	.028	.027	.032	.038	.039
	-.8	.025	.026	.026	.069	.029	.044
	.8	.997	.985	.996	.053	.434	.055
	.5	.573	.445	.559	.024	.069	.039
100	.0	.053	.058	.058	.036	.049	.043
(-2.89)	-.5	.024	.031	.026	.043	.058	.046
	-.8	.028	.035	.028	.062	.078	.050
	.8	.999	.977	.993	.069	.371	.054
	.5	.604	.378	.489	.076	.058	.045
250	.0	.049	.052	.058	.065	.047	.044
(-2.88)	-.5	.024	.039	.035	.063	.048	.047
	-.8	.027	.037	.032	.069	.037	.044
	.8	.999	.961	.984	.153	.403	.058
	.5	.610	.312	.402	.081	.057	.046
500	.0	.053	.054	.058	.069	.052	.046
(-2.87)	-.5	.024	.037	.037	.062	.044	.046
	-.8	.021	.036	.035	.065	.035	.045
	.8	.999	.932	.967	.163	.229	.051
	.5	.624	.254	.332	.096	.056	.050
1000	.0	.049	.050	.055	.069	.049	.047
(-2.86)	-.5	.024	.043	.044	.066	.051	.048
	-.8	.024	.044	.045	.070	.044	.051

NOTE: the proportion of statistics less than the 5% critical value from Fuller[1976, p. 373, Table 8.5.2] for the regression "t-test" for a unit root τ_μ against the alternative hypothesis that the process is stationary around a constant mean. Based on 10,000 replications of an ARIMA(0,1,1) process,

$$(Y_t - Y_{t-1}) = \epsilon_t - \theta\, \epsilon_{t-1},\quad t=1,\ldots T.$$

The Dickey-Fuller critical values are in parentheses under the sample size. The AR(1) test is based on equation (4); the Phillips corrections to the AR(1) test, $Z_{\tau\mu}$, use equations (6) and (7); the ARMA(1,1) test uses equation (3); the AR(ℓ_4) and AR(ℓ_{12}) tests use equation (2) with ℓ_4 and ℓ_{12} lags, respectively, where ℓ_4 and ℓ_{12} are defined in (13a,b). The standard error for these estimates of the size of the tests is .007.

Table 2a. Empirical Size for 1% Level Test

Based on Dickey-Fuller Distribution of τ_τ for $\rho_\tau=1$

Sample Size, T (DF crit. value)	Moving Avg Parameter θ	AR(1)	$Z_{\tau\tau}(\ell_4)$	$Z_{\tau\tau}(\ell_{12})$	ARMA(1,1)	AR(ℓ_4)	AR(ℓ_{12})
	.8	.669	.669	.670	.033	.182	.007
	.5	.241	.241	.251	.041	.046	.007
25	.0	.010	.010	.016	.027	.010	.009
(-4.38)	-.5	.002	.002	.008	.024	.036	.013
	-.8	.002	.002	.008	.030	.049	.015
	.8	.989	.987	.993	.049	.232	.009
	.5	.470	.452	.531	.021	.025	.007
50	.0	.010	.011	.008	.003	.008	.007
(-4.15)	-.5	.001	.002	.002	.007	.006	.008
	-.8	.001	.001	.002	.025	.003	.010
	.8	1.000	.999	1.000	.033	.307	.014
	.5	.612	.537	.703	.003	.020	.007
100	.0	.009	.011	.008	.002	.010	.008
(-4.04)	-.5	.002	.003	.002	.008	.014	.008
	-.8	.002	.002	.001	.015	.027	.008
	.8	1.000	1.000	1.000	.004	.322	.012
	.5	.688	.485	.676	.003	.015	.009
250	.0	.011	.013	.014	.006	.010	.008
(-3.99)	-.5	.002	.004	.002	.009	.008	.009
	-.8	.001	.004	.002	.014	.005	.009
	.8	1.000	.999	1.000	.020	.399	.012
	.5	.709	.386	.575	.016	.016	.009
500	.0	.012	.012	.014	.015	.010	.009
(-3.98)	-.5	.001	.004	.003	.013	.007	.008
	-.8	.001	.004	.003	.016	.005	.009
	.8	1.000	.998	1.000	.067	.169	.013
	.5	.720	.300	.469	.034	.013	.010
1000	.0	.010	.012	.014	.020	.010	.009
(-3.96)	-.5	.002	.007	.006	.020	.010	.010
	-.8	.002	.006	.005	.026	.007	.010

NOTE: the proportion of statistics less than the 1% critical value from Fuller[1976, p. 373, Table 8.5.2] for the regression "t-test" for a unit root τ_τ against the alternative hypothesis that the process is stationary around a time trend. Based on 10,000 replications of an ARIMA(0,1,1) process,

$$(Y_t - Y_{t-1}) = \epsilon_t + \theta\, \epsilon_{t-1},\ t=1,\ldots T.$$

The Dickey-Fuller critical values are in parentheses under the sample size. The AR(1) test is based on equation (8); the Phillips corrections to the AR(1) test, $Z_{\tau\tau}$, use equations (12) and (6); the ARMA(1,1) test uses equation (9); the AR(ℓ_4) and AR(ℓ_{12}) tests use equation (10) with ℓ_4 and ℓ_{12} lags, respectively, where ℓ_4 and ℓ_{12} are defined in (13a,b). The standard error for these estimates of the size of the tests is .001.

Table 2b. Empirical Size for 5% Level Test

Based on Dickey-Fuller Distribution of τ_τ for $\rho_\tau=1$

Sample Size, T (DF crit. value)	Moving Avg Parameter θ	AR(1)	$Z_{\tau\tau}(\ell_4)$	$Z_{\tau\tau}(\ell_{12})$	ARMA(1,1)	AR(ℓ_4)	AR(ℓ_{12})
	.8	.900	.902	.867	.052	.466	.033
	.5	.514	.509	.484	.056	.166	.034
25	.0	.050	.051	.048	.042	.052	.041
(-3.60)	-.5	.013	.013	.022	.043	.120	.047
	-.8	.011	.009	.019	.062	.159	.059
	.8	1.000	.999	1.000	.070	.518	.045
	.5	.709	.669	.753	.033	.099	.032
50	.0	.052	.056	.038	.010	.045	.034
(-3.50)	-.5	.009	.013	.010	.026	.033	.039
	-.8	.009	.010	.009	.058	.020	.044
	.8	1.000	1.000	1.000	.047	.568	.055
	.5	.794	.704	.831	.006	.079	.039
100	.0	.054	.060	.050	.015	.044	.040
(-3.45)	-.5	.011	.020	.011	.031	.061	.040
	-.8	.007	.016	.009	.047	.096	.043
	.8	1.000	1.000	1.000	.009	.551	.056
	.5	.841	.640	.789	.014	.064	.042
250	.0	.051	.062	.065	.032	.050	.047
(-3.43)	-.5	.008	.026	.016	.042	.042	.042
	-.8	.008	.026	.014	.051	.030	.043
	.8	1.000	1.000	1.000	.046	.613	.057
	.5	.853	.545	.704	.041	.065	.046
500	.0	.052	.057	.067	.057	.049	.048
(-3.42)	-.5	.008	.030	.028	.061	.042	.046
	-.8	.007	.027	.026	.063	.029	.048
	.8	1.000	.999	1.000	.100	.350	.051
	.5	.858	.453	.600	.071	.053	.047
1000	.0	.053	.056	.063	.072	.051	.046
(-3.41)	-.5	.008	.036	.037	.069	.048	.049
	-.8	.008	.039	.038	.075	.041	.051

NOTE: the proportion of statistics less than the 5% critical value from Fuller[1976, p. 373, Table 8.5.2] for the regression "t-test" for a unit root τ_τ against the alternative hypothesis that the process is stationary around a time trend. Based on 10,000 replications of an ARIMA(0,1,1) process,

$$(Y_t - Y_{t-1}) = \epsilon_t - \theta\,\epsilon_{t-1},\ t=1,\ldots T.$$

The Dickey-Fuller critical values are in parentheses under the sample size. The AR(1) test is based on equation (8); the Phillips corrections to the AR(1) test, $Z_{\tau\tau}$, use equations (12) and (6); the ARMA(1,1) test uses equation (9); the AR(ℓ_4) and AR(ℓ_{12}) tests use equation (10) with ℓ_4 and ℓ_{12} lags, respectively, where ℓ_4 and ℓ_{12} are defined in (13a,b). The standard error for these estimates of the size of the tests is .007.

Table 4a. Empirical Size for 1% Level Test

Based on Dickey-Fuller Distribution of $T(\hat{\rho}_\tau - 1)$ for $\rho_\tau = 1$

Sample Size, T (DF crit. value)	Moving Avg Parameter θ	AR(1)	$Z_{\rho\tau}(l_4)$	$Z_{\rho\tau}(l_{12})$	ARMA(1,1)	AR(l_4)	AR(l_{12})
	.8	.721	.721	.370	.166	.711	.121
	.5	.267	.259	.156	.145	.306	.122
25	.0	.008	.009	.003	.057	.106	.117
(-22.5)	-.5	.000	.000	.000	.016	.216	.110
	-.8	.000	.000	.000	.010	.279	.112
	.8	.994	.990	.981	.335	.746	.245
	.5	.505	.469	.574	.130	.198	.232
50	.0	.008	.009	.005	.032	.090	.233
(-25.7)	-.5	.000	.001	.000	.014	.060	.212
	-.8	.000	.000	.000	.010	.033	.207
	.8	1.000	.999	1.000	.288	.684	.322
	.5	.649	.541	.745	.064	.113	.260
100	.0	.009	.012	.008	.020	.061	.270
(-27.4)	-.5	.000	.002	.000	.011	.086	.271
	-.8	.000	.001	.000	.009	.126	.280
	.8	1.000	1.000	1.000	.130	.544	.195
	.5	.729	.486	.702	.028	.048	.147
250	.0	.012	.016	.016	.017	.036	.152
(-28.4)	-.5	.000	.003	.000	.009	.028	.149
	-.8	.000	.003	.000	.012	.017	.141
	.8	1.000	.999	1.000	.136	.532	.094
	.5	.748	.385	.596	.026	.034	.074
500	.0	.010	.012	.015	.013	.019	.070
(-28.9)	-.5	.000	.004	.002	.011	.015	.069
	-.8	.000	.004	.002	.010	.008	.072
	.8	1.000	.997	1.000	.102	.237	.043
	.5	.746	.298	.477	.040	.019	.039
1000	.0	.009	.010	.014	.015	.016	.039
(-29.5)	-.5	.000	.006	.004	.015	.014	.040
	-.8	.000	.006	.004	.015	.011	.041

NOTE: the proportion of statistics less than the 1% critical value from Fuller[1976, p. 371, Table 8.5.1] for the normalized bias of the unit root estimator, $T(\hat{\rho}_\tau - 1)$, where a time trend is included as an additional regressor in the estimated model. Based on 10,000 replications of an ARIMA(0,1,1) process,

$$(Y_t - Y_{t-1}) = \epsilon_t + \theta\, \epsilon_{t-1}, \quad t=1,\ldots T.$$

The Dickey-Fuller critical values are in parentheses under the sample size. The AR(1) test is based on equation (8); the Phillips corrections to the AR(1) test, $Z_{\rho\tau}$, use equations (11) and (6); the ARMA(1,1) test uses equation (9); the AR(l_4) and AR(l_{12}) tests use equation (10) with l_4 and l_{12} lags, respectively, where l_4 and l_{12} are defined in (13a,b). The latter tests use Fuller's[1976] correction c multiplied times the raw test statistic, where $c = 1/(1-\phi_1-\ldots-\phi_p)$ is a function of the additional AR parameters estimated for that model. The standard error for these estimates of the size of the tests is .001.

Table 4b. Empirical Size for 5% Level Test

Based on Dickey-Fuller Distribution of $T(\hat{\rho}_\tau-1)$ for $\rho_\tau=1$

Sample Size, T (DF crit. value)	Moving Avg Parameter θ	AR(1)	$Z_{\rho\tau}(l_4)$	$Z_{\rho\tau}(l_{12})$	ARMA(1,1)	AR(l_4)	AR(l_{12})
	.8	.927	.927	.652	.266	.845	.139
	.5	.546	.531	.359	.234	.457	.145
25	.0	.040	.046	.014	.113	.187	.138
(-17.9)	-.5	.003	.007	.001	.058	.332	.132
	-.8	.000	.004	.000	.036	.403	.133
	.8	1.000	.999	.998	.440	.858	.292
	.5	.740	.673	.776	.208	.320	.283
50	.0	.045	.056	.024	.093	.171	.273
(-19.8)	-.5	.002	.010	.001	.055	.121	.259
	-.8	.001	.008	.000	.045	.077	.251
	.8	1.000	1.000	1.000	.376	.821	.408
	.5	.826	.707	.852	.139	.220	.343
100	.0	.050	.061	.045	.071	.143	.355
(-20.7)	-.5	.003	.016	.002	.050	.171	.353
	-.8	.001	.015	.001	.046	.238	.363
	.8	1.000	1.000	1.000	.261	.719	.293
	.5	.869	.641	.805	.089	.128	.238
250	.0	.052	.062	.069	.064	.097	.248
(-21.3)	-.5	.001	.023	.011	.053	.082	.237
	-.8	.001	.025	.011	.049	.055	.231
	.8	1.000	1.000	1.000	.213	.711	.186
	.5	.879	.551	.719	.080	.095	.152
500	.0	.053	.059	.068	.062	.072	.151
(-21.5)	-.5	.002	.030	.025	.054	.062	.147
	-.8	.001	.027	.023	.050	.037	.149
	.8	1.000	.999	1.000	.136	.431	.114
	.5	.879	.455	.611	.081	.070	.099
1000	.0	.052	.055	.063	.059	.065	.107
(-21.8)	-.5	.001	.035	.035	.053	.062	.109
	-.8	.001	.035	.034	.054	.046	.103

NOTE: the proportion of statistics less than the 5% critical value from Fuller[1976, p. 371, Table 8.5.1] for the normalized bias of the unit root estimator, $T(\hat{\rho}_\tau-1)$, where a time trend is included as an additional regressor in the estimated model. Based on 10,000 replications of an ARIMA(0,1,1) process,

$$(Y_t - Y_{t-1}) = \epsilon_t + \theta\, \epsilon_{t-1},\ t=1,\ldots T.$$

The Dickey-Fuller critical values are in parentheses under the sample size. The AR(1) test is based on equation (8); the Phillips corrections to the AR(1) test, $Z_{\rho\tau}$, use equations (11) and (6); the ARMA(1,1) test uses equation (9); the AR(l_4) and AR(l_{12}) tests use equation (10) with l_4 and l_{12} lags, respectively, where l_4 and l_{12} are defined in (13a,b). The latter tests use Fuller's[1976] correction c multiplied times the raw test statistic, where $c=1/(1+\phi_1+\ldots+\phi_p)$ is a function of the additional AR parameters estimated for that model. The standard error for these estimates of the size of the tests is .007.

Table 5a. 5% and 1% Fractiles of the Phillips-Perron Test,

$Z_{\rho\mu}(\ell)$, for an ARIMA(1,0,1) model with ρ_μ=1, θ=.8, and T=1,000 or 10,000

	Sample Size T = 1,000				Sample Size T = 10,000			
	Residuals		Differences		Residuals		Differences	
Lags ℓ	5% Fractile	1% Fractile	5% Fractile	1% Fractile	5% Fractile	1% Fractile	5% Fractile	1% Fractile
0	-366.2 (.999)	-466.9 (.996)			-533.5 (1.00)	-730.7 (.999)		
1	-296.2 (.986)	-401.5 (.958)	-187.4 (.983)	-239.5 (.947)	-301.0 (.993)	-421.8 (.970)	-274.8 (.993)	-372.6 (.969)
2	-299.8 (.967)	-418.9 (.920)	-128.6 (.946)	-162.6 (.875)	-232.7 (.970)	-341.2 (.910)	-187.2 (.970)	-259.3 (.906)
3	-322.6 (.953)	-456.6 (.901)	-98.5 (.903)	-124.7 (.806)	-204.5 (.936)	-310.1 (.852)	-143.6 (.929)	-197.2 (.839)
4	-351.8 (.944)	-498.3 (.893)	-80.1 (.860)	-101.8 (.755)	-194.2 (.902)	-300.7 (.807)	-118.8 (.884)	-160.8 (.779)
ℓ_4	-546.1 (.710)	-817.8 (.672)	-53.2 (.768)	-67.6 (.647)	24.3 (.024)	-97.9 (.022)	-53.7 (.611)	-73.7 (.414)
ℓ_{12}	-953.8 (.976)	-1291.5 (.954)	-24.4 (.632)	-30.0 (.521)	-541.0 (.851)	-975.5 (.784)	-27.0 (.268)	-36.9 (.120)
DF	-14.1 (.050)	-20.7 (.010)	-14.1 (.050)	-20.7 (.010)	-14.1 (.050)	-20.7 (.010)	-14.1 (.050)	-20.7 (.010)

NOTE: the sampling distribution of the normalized bias statistic, $Z_{\rho\mu}(\ell)$, against the alternative hypothesis that the process is stationary around a constant mean. Based on 10,000 replications of an ARIMA(0,1,1) process,

$$(Y_t - Y_{t-1}) = \epsilon_t - \theta\, \epsilon_{t-1}, \qquad t=1,\ldots T,$$

with θ=.8 and T=1,000 or 10,000. The Phillips-Perron corrections use equations (5), (6) and (7) for ℓ lags of the residual autocorrelations. For T=1,000, ℓ_4=7 and ℓ_{12}=21; for T=10,000, ℓ_4=12 and ℓ_{12}=37. The percentage of rejections using the Dickey-Fuller critical value is in parentheses under each of the fractiles (i.e., for a 5% level test, this should be .05 if the approximation to the Dickey-Fuller distribution is accurate). The last row, labeled DF, contains the asymptotic Dickey-Fuller critical values and rejection percentages.

Table 5b. 5% and 1% Fractiles of the Phillips-Perron Test,

$Z_{\tau\mu}(\ell)$, for an ARIMA(1,0,1) model with ρ_μ=1, θ=.8, and T=1,000 or 10,000

	Sample Size T = 1,000				Sample Size T = 10,000			
	Residuals		Differences		Residuals		Differences	
Lags ℓ	5% Fractile	1% Fractile	5% Fractile	1% Fractile	5% Fractile	1% Fractile	5% Fractile	1% Fractil
0	-14.96 (1.00)	-17.43 (.997)			-16.56 (1.00)	-19.51 (.997)		
1	-13.82 (.992)	-16.55 (.968)	-11.68 (.991)	-13.89 (.960)	-12.60 (.986)	-15.00 (.953)	-12.06 (.986)	-14.15 (.951)
2	-13.88 (.976)	-16.79 (.933)	-10.41 (.960)	-12.63 (.888)	-11.12 (.949)	-13.60 (.879)	-10.05 (.945)	-11.99 (.873)
3	-14.25 (.962)	-17.28 (.911)	-9.79 (.915)	-12.08 (.804)	-10.50 (.909)	-13.03 (.825)	-8.92 (.898)	-10.62 (.805)
4	-14.73 (.955)	-17.85 (.902)	-9.38 (.863)	-11.71 (.727)	-10.25 (.868)	-12.89 (.778)	-8.19 (.851)	-9.73 (.736)
ℓ_4	-17.53 (.674)	-21.46 (.644)	-8.91 (.716)	-11.45 (.528)	-3.28 (.055)	-8.75 (.048)	-5.87 (.577)	-7.20 (.409)
ℓ_{12}	-22.64 (.980)	-26.53 (.960)	-9.00 (.306)	-12.31 (.107)	-16.69 (.835)	-22.40 (.767)	-4.68 (.329)	-5.96 (.188)
DF	-2.86 (.050)	-3.43 (.010)	-2.86 (.050)	-3.43 (.010)	-2.86 (.050)	-3.43 (.010)	-2.86 (.050)	-3.43 (.010)

NOTE: the sampling distribution of the normalized bias statistic, $Z_{\tau\mu}(\ell)$, against the alternative hypothesis that the process is stationary around a constant mean. Based on 10,000 replications of an ARIMA(0,1,1) process,

$$(Y_t - Y_{t-1}) = \epsilon_t - \theta\, \epsilon_{t-1}, \qquad t=1,\ldots T,$$

with θ=.8 and T=1,000 or 10,000. The Phillips-Perron corrections use equations (5), (6) and (7) for ℓ lags of the residual autocorrelations. For T=1,000, ℓ_4=7 and ℓ_{12}=21; for T=10,000, ℓ_4=12 and ℓ_{12}=37. The percentage of rejections using the Dickey-Fuller critica value is in parentheses under each of the fractiles (i.e., for a 5% level test, this should be .05 if the approximation to the Dickey-Fuller distribution is accurate). The last row, labeled DF, contains the asymptotic Dickey-Fuller critical values and rejection percentages.