EXAMINATION OF SOME MORE POWERFUL MODIFICATIONS OF THE DICKEY-FULLER TEST

BY STEPHEN LEYBOURNE, TAE-HWAN KIM AND PAUL NEWBOLD

University of Nottingham

Abstract. Although the $t-$ratio variant of the Dickey-Fuller test is the most commonly applied unit root test in practical applications, it has been known for some time that readily implementable, more powerful modifications are available. We explore the large sample properties of five of these modified tests, and the small sample properties of these five plus six hybrids. As a result of this study we recommend two particular test procedures.

Keywords. Dickey-Fuller test; MAX test; weighted symmetric estimation; recursive and GLS detrending; power comparison.

1. INTRODUCTION

Very often, in the analysis of economic time series, a preliminary step is to test the null hypothesis that an individual series of T observations is integrated of order one, $I(1)$, against the alternative that it is integrated of order zero, $I(0)$. The most commonly applied test of this sort is the $t-$ratio (pivotal statistic) variant of the Dickey-Fuller test of Dickey and Fuller (1979). In its most basic variant, the test statistic, which has a non-standard limiting null distribution, is the $t-$ratio, DF, associated with the OLS estimator of $(\rho - 1)$ in the model

$$y_t = \gamma' z_t + \rho y_{t-1} + \epsilon_t; \qquad t = 2, 3, ..., T \tag{1}$$

with ϵ_t taken to be independent and identically distributed with mean zero, and variance σ^2, and where either $z_t = 1$ or $z_t = [1, t]'$ and γ is a conformable vector of unknown parameters. In the former case the test is of a driftless random walk against a stationary first order autoregression with unknown mean while in the latter it is of a random walk with drift against an unknown linear trend with stationary first order autoregressive errors. The test is easily extended to allow higher order autoregressive processes (which might be viewed as approximations to more general processes) through augmentation of (1) by lagged first differences of y_t, a consideration that accounts for its popularity.

It has been known for several years that, with a modest amount of computational effort, more powerful modifications of this test are available. Our purpose here is to compare some of these tests. The DF test is asymptotically equivalent to prior OLS detrending of y_t, followed by the fitting to the residuals $\tilde{y}_t$ of a model of the form (1) with $\gamma = 0$. Three modifications involve alternative detrending, motivated by derivations of the asymptotic Gaussian power envelope:

1. Elliott *et al* (1996) apply generalised least squares de-trending, taking $\tilde{y}_t$ as the residuals from the regression of $[y_1, y_2 - \alpha y_1, ..., y_T - \alpha y_{T-1}]'$ on $[z_1, z_2 - \alpha z_1, ..., z_T - \alpha z_{T-1}]'$, where $\alpha = 1 + \bar{c}T^{-1}$, with $\bar{c} \in (-\infty, 0)$ a constant specified from consideration of the power envelope. The resultant test, which we denote GLS, is based on the fitting to $\tilde{y}_t$ of a model of the form (1) with $\gamma = 0$.

2. Elliott (1999) notes that the above test is motivated by an alternative model in which the initial observation is taken to be fixed, while a frequently more attractive assumption is of full covariance stationarity, so that in terms of (1) the initial deviation from trend is a zero-mean random variable with variance $\sigma^2(1 - \rho^2)^{-1}$. The initial GLS detrending would then generate $\tilde{y}_t$ as the residuals from the regression of $[(1 - \alpha^2)^{1/2} y_1, y_2 - \alpha y_1, ..., y_T -$

$\alpha y_{T-1}]'$ on $[(1-\alpha^2)^{1/2}z_1, z_2 - \alpha z_1, ..., z_T - \alpha z_{T-1}]'$. This approach is often termed "unconditional," and we denote the resulting unit root test statistic GLS_u. Several proposals for the choice of $\bar{c}$ in GLS and GLS_u have been made. Elliott *et al* (1996) suggest $\bar{c} = -7$ for $z_t = 1$ and $\bar{c} = -13.5$ for $z_t = [1, t]'$; Elliott (1999) suggests $\bar{c} = -10$ in both cases. We adopt these recommendations in our analysis of the tests.

3. Taylor (2002) considers recursive OLS detrending, a proposal which has the advantage of not requiring the somewhat arbitrary specification of a constant parameter. Thus $\tilde{y}_t$ are the residuals from the OLS regression of y_j on z_j, $j \leq t$. Again, the unit root test statistic, which we denote REC, follows from the regression (1) with $\tilde{y}_t$ in place of y_t and $\gamma = 0$.

Two previously proposed approaches retain OLS detrending (explicitly or implicitly) and are motivated by the fact that, in the Gaussian case, under stationarity, forward- and backward-looking finite order AR models have identical covariance structures. These are:

4. Pantula *et al* (1994) first employ OLS detrending to generate residuals $\tilde{y}_t$. They then recommend a test based on weighted symmetric estimation of ρ, through the minimization of

$$Q(\rho) = \sum_{t=2}^{T} w_t(\tilde{y}_t - \rho\tilde{y}_{t-1})^2 + \sum_{t=1}^{T-1}(1 - w_{t+1})(\tilde{y}_t - \rho\tilde{y}_{t+1})^2 \quad ; \quad w_t = T^{-1}(t-1)$$

from which a pivotal statistic readily follows. We denote this statistic WS.

5. Leybourne (1995) proposes OLS estimation of (1), together with OLS estimation of the corresponding model for the reversed series; that is

$$v_t = \delta' z_t + \rho v_{t-1} + \eta_t; \qquad t = 2, 3, ..., T \tag{2}$$

where $v_t = y_{T+1-t}$. Denote by DF_f the Dickey-Fuller $t-$ratio from (1) and by DF_r the corresponding statistic from (2). Leybourne's proposed statistic, which we denote MAX, is then $\max(DF_f, DF_r)$.

The limiting null distributions of these five modified tests are all given in the cited literature. As regards the alternative hypothesis, we strongly prefer true stationarity, and restrict attention to this case in the remainder of the paper, in line with the view of Pantula *et al* (1994), quoted approvingly by Taylor (2002), that formulations such as that where the deviation from trend of the first observation has the same variance as the error terms might reasonably be assumed in "a modest number of situations." We therefore consider power under the more natural stationarity alternative, where the deviation from trend of the first observation has variance $\sigma^2(1-\rho^2)^{-1}$. One possible procedure is through calculation of the local asymptotic power, which follows directly from the limiting distribution of the test statistics, where as in Chan and Wei (1987), Phillips and Perron (1988) and elsewhere we set $\rho = 1 + cT^{-1}$, for fixed c, using the "local-to-unity asymptotics" approach. This limiting distribution is given for the MAX test in Section 2. Those for the other tests are given by, or can be directly inferred from, Elliott *et al* (1996), Elliott (1999), and Taylor (2002). We go on to compare the results with the Gaussian power envelope, given by Elliott (1999).

It emerges from Section 2 that some at least of the tests have local asymptotic power close to the envelope. However, this is certainly insufficient to conclude that the behaviour of the tests, as they are applied in practice, with finite sample sizes, will be the same. Accordingly, in Section 3 we report results of a simulation study on the finite sample power and size properties of the tests. As well as the standard Dickey-Fuller test and the five modifications noted earlier, we consider also six "hybrid" tests, in which each detrending procedure is used in conjunction with both the WS approach and the MAX approach. For example, the statistic we denote GLS^{MAX} first applies GLS detrending, and then applies the MAX principle, the test statistic being the maximum of the $t-$ratios for testing $\rho = 1$ from the estimation of (1) with $\gamma = 0$ and (2) with $\delta = 0$.

2. ASYMPTOTIC DISTRIBUTION OF THE MAX STATISTICS UNDER THE LOCAL ALTERNATIVE

Suppose that the time series y_t is generated through

$$\begin{aligned} y_t &= \rho y_{t-1} + \epsilon_t \\ \rho &= 1 + \frac{c}{T} \end{aligned} \tag{3}$$

where $c \in (-\infty, 0)$. We impose the following assumptions on the initial value y_1 and the error term ϵ_t.

Assumption 1. (i) y_1 is distributed with mean zero and variance $\sigma^2(1-\rho^2)^{-1}$, (ii) ϵ_t is i.i.d.$(0, \sigma^2)$ and (iii) y_1 is uncorrelated with ϵ_t, $t \geq 2$.

Assumption (i) is adopted from Elliott (1999), implying that the first observation y_1 is drawn from the unconditional distribution of y_t. Assumption (ii) is made for clarity and simplicity and it can be relaxed to allow ϵ_t to be a martingale difference sequence (see, for instance, Banerjee *et al*, 1992). The DF regression based on the forward series y_t is

$$\Delta y_t = \hat{\gamma}' z_t + \hat{\rho} y_{t-1} + \hat{\epsilon}_t \tag{4}$$

while the DF regression based on the reverse series $v_t = y_{T+1-t}$ is

$$\Delta v_t = \tilde{\gamma}' z_t + \tilde{\rho} v_{t-1} + \tilde{\eta}_t. \tag{5}$$

Let DF_f denote the Dickey-Fuller $t-$ratio from (4) and DF_r denote the corresponding statistic from (5). The MAX statistic is then $MAX = \max(DF_f, DF_r)$. The following theorem gives the limiting distribution of the MAX test statistics under the local alternative.

Theorem 1 *If y_t is generated by (3) and Assumption 1 holds, then;*

(a) If the fitted model contains a constant only, so that in (4) and (5) $z_t = 1$,

$$MAX \Rightarrow \max(F_0, R_0)$$

where

$$F_0 = \frac{0.5\{J_c(1)^2 - 1\} - H_c J_c(1)}{(G_c - H_c^2)^{1/2}}$$

$$R_0 = \frac{-0.5\{J_c(1)^2 + 1\} + H_c J_c(1)}{(G_c - H_c^2)^{1/2}}$$

and

$$J_c(r) = W_c(r) + (e^{rc} - 1) Z_c$$

$$H_c = \int_0^1 J_c(r) dr$$

$$G_c = \int_0^1 J_c(r)^2 dr.$$

Here, $W_c(r)$ is an Ornstein-Uhlenbeck process defined as $W_c(r) = c \int_0^r e^{c(r-\lambda)} W(\lambda) d\lambda + W(r)$, $W(r)$ is a standard Brownian motion process defined as the limit of $\sigma^{-1} T^{-1/2} \sum_{t=1}^{rT} \epsilon_t$ and Z_c is a random variable with mean zero and variance $(-2c)^{-1}$.

(b) If the fitted model contains a linear trend, so that in (4) and (5) $z_t = [1, t]'$,

$$MAX \Rightarrow \max(F_1, R_1)$$

where

$$F_1 = \frac{0.5\{J_c(1)^2 - 1\} - 6M_c J_c(1) + 2H_c J_c(1) + 12 H_c M_c - 6H_c^2}{(G_c - 12M_c^2 + 12 H_c M_c - 4H_c^2)^{1/2}}$$

$$R_1 = \frac{-0.5\{J_c(1)^2 + 1\} + 6M_c J_c(1) - 2H_c J_c(1) - 12 H_c M_c + 6H_c^2}{(G_c - 12M_c^2 + 12 H_c M_c - 4H_c^2)^{1/2}}$$

and

$$M_c = \int_0^1 r J_c(r) dr.$$

The asymptotic distributions of the MAX statistics under the unit root hypothesis $\rho = 1$ can be shown as a special case for $c = 0$. The variance of $(e^{rc} - 1)Z_c$ is given by $(e^{rc} - 1)^2(-2c)^{-1}$ and it can be shown to converge to zero as $c \rightarrow 0$ by applying the L'hopital's rule. Also $W_c(r)$ converges to the standard Brownian motion process $W(r)$ as $c \rightarrow 0$. Hence, we have the following results: $J_c(1), H_c, G_c$ and M_c converge to $W(1), H = \int_0^1 W(r)dr, G = \int_0^1 W(r)^2 dr$ and $M = \int_0^1 rW(r)dr$ respectively as $c \rightarrow 0$. Therefore, when the fitted model contains a constant only, the limiting null distribution is given by

$$MAX \Rightarrow \max(F_{n0}, R_{n0})$$

where

$$\begin{aligned} F_{n0} &= \frac{0.5\{W(1)^2 - 1\} - HW(1)}{(G - H^2)^{1/2}} \\ R_{n0} &= \frac{-0.5\{W(1)^2 + 1\} + HW(1)}{(G - H^2)^{1/2}}. \end{aligned}$$

(this result is given in Leybourne (1995)). When the fitted model contains both a constant and a linear trend, the limiting null distribution is now given by

$$MAX \Rightarrow \max(F_{n1}, R_{n1})$$

where

$$\begin{aligned} F_{n1} &= \frac{0.5\{W(1)^2 - 1\} - 6MW(1) + 2HW(1) + 12HM - 6H^2}{(G - 12M^2 + 12HM - 4H^2)^{1/2}} \\ R_{n1} &= \frac{-0.5\{W(1)^2 + 1\} + 6MW(1) - 2HW(1) - 12HM + 6H^2}{(G - 12M^2 + 12HM - 4H^2)^{1/2}}. \end{aligned}$$

Note that it is readily shown that the same limiting null distributions arise if ϵ_t follows a stationary $AR(p^*)$ process with martingale difference disturbances, provided that (4) and (5) are augmented with $p \geq p^*$ lagged changes in y_t and v_t respectively (for details see Leybourne *et al,* 2002).

Given the local asymptotic distributions of the various test statistics, and the critical values following from the asymptotic null distributions, asymptotic local power can be calculated. Results for 0.05-level tests are shown in Table 1, where ENV denotes the asymptotic Gaussian power envelope, taken from Elliott (1999). These results were obtained by simulating 50000 replications of the appropriate limiting functionals, using series of 5000 Gaussian white noise innovations. Here and throughout, all calculations were programmed in GAUSS. Notice first that, with the exception of the constant case for $c = -20, -25$ where GLS is inferior, DF is outperformed in terms of asymptotic local power by all five modified tests in both the constant and linear trend cases. It also emerges that the REC, WS, and MAX tests have asymptotic local powers on or very close to the envelope for all the values of c considered. This is not, however, the situation for the two tests based on GLS detrending; this being particularly evident in the constant only case. As previously noted by Elliott (1999) and Taylor (2002), the relatively poor performance of the GLS_u test is rather ironic, given it was designed with the strictly stationary alternative analysed here in mind.

3. FINITE SAMPLE SIMULATIONS

The augmented version of the DF test is based on fitting the model

$$y_t = \gamma' z_t + \rho y_{t-1} + \sum_{j=1}^{p} \phi_j \Delta y_{t-j} + \epsilon_t \qquad (6)$$

with corresponding elaborations of the modified tests. These modifications are all designed to increase power, and the results of the previous section show that, in very large samples, substantial power gains can be achieved. In this section we assess the possibility of achievable power gains for sample sizes of practical interest. To some extent, such gains have been previously demonstrated for these modified tests, though there has been little exploration of their hybrids. We consider the case where it is known that $p = 0$ in (6), and also the more realistic situation where p is unknown and is selected through a data-dependent rule. The tests' size in the latter case is also considered, allowing for additional autoregressive and moving average behaviour.

We generated data from the model (3) under Assumption 1 with $\sigma^2 = 1$, y_1 and ϵ_t normally distributed, for $T = 75, 150$. Then, we constructed the DF test, its five modifications, and the six hybrid tests based on those modifications tests from fitting (6) with $p = 0$. Table 2 shows the empirical power for nominal 0.05-level tests, constant only and linear trend cases; finite sample null critical values for the tests having being calculated by setting $c = 0$ and $y_1 = 0$. Here and throughout the remainder of the paper results are based on 20000 replications. The rankings for DF and its five modifications broadly mimic the asymptotic ones given in Table 1.[1] The

[1]Throughout Tables 2-5 we establish rankings of tests based on informal comparisons of the tests' performance. A more rigorous approach would be to calculate, for any two point estimates (cell entries) $\hat{p}_1$ and $\hat{p}_2$, the $t-$statistic $(\hat{p}_1 - \hat{p}_2)/s.e.(\hat{p}_1 - \hat{p}_2)$ where

$$s.e.(\hat{p}_1 - \hat{p}_2) = \{n^{-1}\hat{p}_1(1 - \hat{p}_1) + n^{-1}\hat{p}_2(1 - \hat{p}_2)\}^{1/2}$$

and n is the number of Monte Carlo replications, to test whether two true rejection probabilities p_1 and p_2 are different. Given $n = 20000$, we find that *any two* of our point estimates which differ by 0.02 or more are significantly different at the 0.05-level of an approximating normal distribution. Hence, we can make a statistically meaningful comparison of two tests' relative performance whenever their entries differ by at least 0.02.

only noticeable departure from this pattern is that in the constant only case GLS performs more competitively when $T = 75$ than in the asymptotic case, particularly in relation to the DF test. The REC, WS, and MAX tests again behave very similarly to each other and emerge as the clearly dominant trio, particularly in the constant only case. Of the six hybrid tests, GLS^{WS} and GLS^{MAX} are dominated by the other four, whilst these four all perform very similarly to REC, WS, and MAX. The only obvious virtue of the hybrid tests is that the two hybrid variants GLS_u perform rather better than standard GLS_u. That no hybrid variant dominates REC, WS, and MAX is perhaps, however, not surprising, given that the near-optimal large sample properties of these three would imply at best only very limited scope for improvement.[2]

Using the same data generating model, Table 3 compares the tests' empirical power (at the nominal 0.05-level) when p in (6) is determined by the data-dependent rule suggested by Ng and Perron (1995); that is downwards testing of lagged difference terms at the 0.10-level, starting from $p_{\max} = 4$. Here, in computing the augmented MAX statistic, the same value of p was used in the forward and reverse regressions, and selected from the forward regression alone. The augmented variant of the WS test is specified in Pantula *et al* (1994). As we might expect, the powers of all the tests are rather lower than their $p = 0$ counterparts of Table 2. More interestingly, however, is that tests based on weighted symmetric estimation generally seem to perform more poorly in this downward testing environment than in the previous fixed lag case, particularly with $T = 75$. Of the non-hybrid tests the pair REC and MAX now appear quite clearly dominant, with the latter showing a small advantage over the former. GLS_u^{MAX} and REC^{MAX} appear dominant among the hybrid tests. Again, however, there would seem no evidence to suggest that either hybrid test GLS_u^{MAX} or REC^{MAX} should be preferred to the standard MAX test.

Although the results of Table 3 suggest a preference for the REC and MAX tests on the basis of empirical power, it is also important to check for size robustness in more elaborate cases. We consider the $ARMA(1, 1, 1)$ generating model

$$(1 - \phi L)\Delta y_t = (1 - \theta L)\epsilon_t.$$

Again, the tests are based on fitted models whose lag length is determined by downwards testing from $p_{\max} = 4$. Table 4 shows the empirical sizes of the tests (nominal 0.05-level) for various choices of ϕ and θ. Generally speaking, the empirical sizes are close to nominal sizes. The exception is the case where $\theta = 0.5$ in which case all the tests are quite badly over-sized, particularly in the linear trend case when $T = 75$. Here, the WS test and

[2]This concurs with Shin and So (2001) who also found that hybrid tests which applied their form of recursive demeaning to near-optimal tests yielded no finite sample power gains.

its hybrids, suffer slightly less from over-sizing, though there is very little to choose between any of the tests once $T = 150$.

Finally, to check size and power robustness to non-normality, we generated data from the model (3) under Assumption 1 with y_1 and ϵ_t generated from a highly skewed distribution, $\chi^2(1) - 1$, and also from a heavy-tailed distribution, Students' t with five degrees of freedom, $t(5)$, (setting $c = 0$ and $y_1 = 0$ for the null case). Tests were based on fitting (6) with $p = 0$, using finite sample critical values derived under the normality assumption. Table 5 shows the empirical size and power of nominal 0.05-level tests, in the constant only case with $T = 75$. The empirical sizes are all very close their nominal value. Under the alternatives, as c deviates from zero, for both the χ^2 and t disturbances, the picture remains almost identical to the corresponding situation for normal disturbances seen in Table 2, with REC, WS, and MAX emergent as the dominant trio. Thus, the possibility of non-normality in the disturbance terms would not seem to a pertinent issue in determining the choice of test.

4. CONCLUSIONS

It is now well known that the $t-$ratio variant of the Dickey-Fuller test has inferior power compared with some quite easily implemented modifications. We have analysed the performance of the DF test and eleven such modifications. Three of these are based on alternatives to OLS detrending prior to fitting the usual DF regression, without intercept or trend, to the residuals. Two others retain, at least implicitly, OLS detrending, but exploit the coincidence of covariance structures of forward- and backward-looking finite order stationary AR models under Gaussianity. Finally, six hybrid tests result from applying the principles of this last pair to the residuals from the three alternative detrending procedures. We have explored both asymptotic and finite sample properties of all the tests. On the basis of our results, for practical application we would recommend two tests in particular - either the REC test of Taylor (2002) or the MAX test of Leybourne (1995). Our power and size simulations suggest there is generally very little to choose between thee two, and there seems no material advantage to be gained in combining the two tests together in hybrid fashion. The REC and MAX tests also have the advantage that they are relatively straightforward to compute (particularly the latter) and, unlike GLS-based tests, do not depend on the choice of user-supplied parameters.

APPENDIX

PROOF OF THEOREM 1. First we prove the constant only case. Note that $y_{rT} = \sum_{s=2}^{rT} \rho^{rT-s}\epsilon_s + \rho^{rT-1}y_1$ and let $\tilde{y}_t = y_t - y_1$. Since $T^{-1/2}\sum_{s=2}^{rT} \rho^{rT-s}\epsilon_s \Rightarrow \sigma W_c(r)$ [by Phillips (1987)], $\rho^{rT} \rightarrow e^{rc}$, and $T^{-1/2}y_1 \Rightarrow \sigma Z_c$, we have $T^{-1/2}\tilde{y}_{rT} \Rightarrow \sigma\{W_c(r) + (e^{rc}-1)Z_c\} = \sigma J_c(r)$. By the continuous mapping theorem, we can immediately obtain $T^{-3/2}\sum_{t=2}^{T} \tilde{y}_{t-1} \Rightarrow \sigma\int_0^1 J_c(r)dr = \sigma H_c$ and $T^{-2}\sum_{t=2}^{T}\tilde{y}_{t-1}^2 \Rightarrow \sigma^2\int_0^1 J_c^2(r)dr = \sigma^2 G_c$. Next consider

$$\begin{aligned} T^{-1}\sum_{2}^{T}\tilde{y}_{t-1}\epsilon_t &= (2\rho)^{-1}(T^{-1}\tilde{y}_T^2 - T^{-1}\sum_{t=2}^{T}\epsilon_t^2 - 2cT^{-2}\sum_{t=2}^{T}\tilde{y}_{t-1}^2 \\ &\quad -2cT^{-1/2}y_1T^{-3/2}\sum_{t=2}^{T}\tilde{y}_{t-1}) + o_p(1) \\ &\Rightarrow \sigma^2\{0.5(J_c(1)^2-1) - c(G_c + Z_cH_c)\} = \sigma^2 E_c. \end{aligned}$$

From (4) we have

$$\begin{aligned} T(\hat{\rho}-\rho) &= (0,1)\begin{bmatrix} 1 & T^{-3/2}\sum_{t=2}^{T}\tilde{y}_{t-1} \\ T^{-3/2}\sum_{t=2}^{T}\tilde{y}_{t-1} & T^{-2}\sum_{t=2}^{T}\tilde{y}_{t-1}^2 \end{bmatrix}^{-1}\begin{bmatrix} T^{-1/2}\sum_{t=2}^{T}\epsilon_t \\ T^{-1}\sum_{t=2}^{T}\tilde{y}_{t-1}\epsilon_t \end{bmatrix} + o_p(1) \\ &\Rightarrow (0,1)A_f^{-1}B_f \end{aligned}$$

where

$$A_f = \begin{bmatrix} 1 & \sigma H_c \\ \sigma H_c & \sigma^2 G_c \end{bmatrix}, \quad B_f = \begin{bmatrix} \sigma W(1) \\ \sigma^2 E_c \end{bmatrix}.$$

Using the fact that $T^{-1}\sum_{t=2}^{T}\hat{\epsilon}_t^2 \xrightarrow{p} \sigma^2$, it can be shown that $(\hat{\rho}-\rho)v\hat{a}r(\hat{\rho})^{-1/2} \Rightarrow (0,1)A_f^{-1}B_f\sigma^{-1}\{(0,1)A_f^{-1}(0,1)'\}^{-1/2}$. Note that $DF_f = c\{Tv\hat{a}r(\hat{\rho})\}^{-1/2} + (\hat{\rho}-\rho)v\hat{a}r(\hat{\rho})^{-1/2} \Rightarrow c(G_c - H_c^2)^{1/2} + \{E_c - H_cW(1)\}(G_c - H_c^2)^{-1/2}$ which simplifies to the result in the theorem.

Now we consider the reverse regression (5). Given that the reverse data generating process is

$$v_t = \rho v_{t-1} + \eta_t$$

where $\eta_t = (1-\rho^2)y_{T+1-t} - \rho\epsilon_{T+2-t}$, we have

$$T(\tilde{\rho}-\rho) = (0,1)\begin{bmatrix} 1 & T^{-3/2}\sum_{t=2}^{T}\tilde{v}_{t-1} \\ T^{-3/2}\sum_{t=2}^{T}\tilde{v}_{t-1} & T^{-2}\sum_{t=2}^{T}\tilde{v}_{t-1}^2 \end{bmatrix}^{-1}\begin{bmatrix} T^{-1/2}\sum_{t=2}^{T}\eta_t \\ T^{-1}\sum_{t=2}^{T}\tilde{v}_{t-1}\eta_t \end{bmatrix}$$

where $\tilde{v}_t = v_t - y_1$. It can be shown that $T^{-3/2}\sum_{t=2}^{T}\tilde{v}_{t-1} \Rightarrow \sigma H_c$, $T^{-2}\sum_{t=2}^{T}\tilde{v}_{t-1}^2 \Rightarrow \sigma^2 G_c$, $T^{-1/2}\sum_{t=2}^{T}\eta_t \Rightarrow -\sigma\{W(1) + 2c(H_c + Z_c)\}$ and $T^{-1}\sum_{t=2}^{T}\tilde{v}_{t-1}\eta_t \Rightarrow$

$-\sigma^2\{E_c + 1 + 2c(G_c + Z_c H_c)\}$. Hence, we have $T(\tilde{\rho} - \rho) \Rightarrow (0,1)A_r^{-1}B_r$ where

$$A_r = A_f, \quad B_r = \begin{bmatrix} -\sigma\{W(1) + 2c(H_c + Z_c)\} \\ -\sigma^2\{E_c + 1 + 2c(G_c + Z_c H_c)\} \end{bmatrix}. \tag{7}$$

The above result together with $T^{-1}\sum_{t=2}^{T}\tilde{\eta}_t^2 \xrightarrow{p} \sigma^2$ implies that $(\tilde{\rho}-\rho)v\hat{a}r(\tilde{\rho})^{-1/2} \Rightarrow (0,1)A_r^{-1}B_r\sigma^{-1}\{(0,1)A_r^{-1}(0,1)'\}^{-1/2}$. Note that $DF_r = c\{Tv\hat{a}r(\tilde{\rho})\}^{-1/2} + (\tilde{\rho}-\rho)v\hat{a}r(\tilde{\rho})^{-1/2} \Rightarrow c(G_c - H_c^2)^{1/2} + \{-E_c - 1 - 2cG_c + H_c(W(1) + 2cH_c)\}(G_c - H_c^2)^{-1/2}$ which simplifies to the result in the theorem. Once we obtain these two results, then we have $MAX = \max(DF_f, DF_r) \Rightarrow \max(F_0, R_0)$ by the continuous mapping theorem.

We now turn to the trend case. Since the proof is very similar to the constant only case, its detail is not presented. The only additional part is to establish the following limits:

$$\begin{aligned}
T^{-5/2}\sum_{t=2}^{T} t\tilde{y}_{t-1} &\Rightarrow \sigma\int_0^1 rJ_c(r)dr = \sigma M_c \\
T^{-3/2}\sum_{t=2}^{T} t\epsilon_t &= T^{-1/2}\tilde{y}_T - T^{-3/2}\sum_{t=2}^{T}\tilde{y}_{t-1} - cT^{-5/2}\sum_{t=2}^{T} t\tilde{y}_{t-1} \\
&\quad -0.5cT^{-1/2}y_1 + o_p(1) \\
&\Rightarrow \sigma\{J_c(1) - H_c - c(M_c + 0.5Z_c)\} \\
T^{-5/2}\sum_{t=2}^{T} t\tilde{v}_{t-1} &= T^{-3/2}\sum_{t=1}^{T-1}\tilde{y}_t - T^{-5/2}\sum_{t=1}^{T-1} t\tilde{y}_t + o_p(1) \\
&\Rightarrow -\sigma(M_c - H_c) \\
T^{-3/2}\sum_{t=2}^{T} t\eta_t &= -2cT^{-3/2}\sum_{t=1}^{T-1} y_t + 2cT^{-5/2}\sum_{t=1}^{T-1} ty_t - \rho T^{-1/2}\sum_{t=2}^{T}\epsilon_{t-1} \\
&\quad +\rho T^{-3/2}\sum_{t=2}^{T} t\epsilon_{t-1} + o_p(1) \\
&\Rightarrow \sigma\{J_c(1) - (1+2c)H_c + cM_c - W(1) - 1.5cZ_c\}.
\end{aligned}$$

References

[1] BANERJEE, A., LUMSDAINE, R.L. and STOCK, J.H. (1992) Recursive and sequential tests of the unit-root and trend-break hypotheses: theory and international evidence. *Journal of Business & Economic Statistics* 10, 271-287.

[2] CHAN, N.H. and WEI, C.Z. (1987) Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics* 15, 1050-1063.

[3] DICKEY, D.A. and FULLER, W.A. (1979) Distribution of estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427-431.

[4] ELLIOTT, G. (1999) Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. *International Economic Review* 40, 767-783.

[5] ELLIOTT, G., ROTHENBERG, T.J. and STOCK, J.H. (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813-836.

[6] LEYBOURNE, S.J. (1995) Testing for unit roots using forward and reverse Dickey-Fuller regressions. *Oxford Bulletin of Economics and Statistics* 57, 559-571.

[7] LEYBOURNE, S.J, KIM T.H. and NEWBOLD, P. (2002) Examination of some more powerful modifications of the Dickey-Fuller test. Mimeo, University of Nottingham.

[8] NG, S. and PERRON, P. (1995) Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association* 90, 268-281.

[9] PANTULA, S.G., FARIAS-GONZALES, G. and FULLER, W.A. (1994) A comparison of unit root test criteria. *Journal of Business & Economic Statistics* 12, 449-459.

[10] PHILLIPS, P.C.B. (1987) Towards a unified asymptotic theory for autoregression. *Biometrika* 74, 535-547.

[11] PHILLIPS, P.C.B. and PERRON, P. (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335-346.

[12] SHIN, D.W. and SO, B.S. (2001) Recursive mean adjustment for unit root tests. *Journal of Time Series Analysis* 22, 595-612.

[13] TAYLOR, A.M.R. (2002) Regression-based unit root tests with recursive mean adjustment for seasonal and non-seasonal time series. *Journal of Business & Economic Statistics* 20, 269-281.

Table 1. Asymptotic local power at nominal 0.05-level.

	constant case					trend case				
c	-5	-10	-15	-20	-25	-5	-10	-15	-20	-25
ENV	.20	.52	.83	.97	1.0	.10	.24	.49	.74	.91
DF	.13	.33	.62	.86	.97	.09	.19	.38	.62	.83
GLS	.19	.44	.63	.76	.84	.10	.24	.46	.68	.83
GLS_u	.15	.38	.69	.90	.98	.10	.24	.47	.73	.91
REC	.19	.50	.81	.96	1.0	.10	.24	.49	.74	.91
WS	.20	.51	.83	.96	1.0	.10	.24	.49	.74	.91
MAX	.20	.50	.82	.96	1.0	.10	.24	.49	.74	.91

Table 2. Finite sample power at nominal 0.05-level, $p = 0$.

$T = 75$.

	constant case					trend case				
c	-5	-10	-15	-20	-25	-5	-10	-15	-20	-25
DF	.13	.34	.66	.89	.98	.09	.20	.41	.68	.88
GLS	.19	.50	.78	.92	.97	.10	.25	.51	.78	.93
GLS_u	.15	.40	.73	.93	.99	.10	.24	.49	.76	.92
REC	.19	.51	.83	.97	1.0	.10	.25	.51	.78	.94
WS	.20	.52	.84	.97	1.0	.10	.25	.51	.78	.94
MAX	.19	.51	.84	.97	1.0	.10	.25	.51	.78	.94
GLS^{WS}	.19	.50	.77	.90	.96	.10	.25	.51	.77	.92
GLS_u^{WS}	.20	.52	.84	.97	1.0	.10	.25	.51	.78	.94
REC^{WS}	.19	.51	.83	.97	1.0	.10	.25	.51	.78	.93
GLS^{MAX}	.19	.50	.77	.91	.97	.10	.25	.52	.78	.93
GLS_u^{MAX}	.19	.51	.84	.97	1.0	.10	.25	.51	.79	.94
REC^{MAX}	.19	.51	.84	.97	1.0	.10	.25	.51	.78	.94

$T = 150$.

	constant case					trend case				
c	-5	-10	-15	-20	-25	-5	-10	-15	-20	-25
DF	.12	.33	.63	.87	.97	.09	.19	.39	.65	.85
GLS	.19	.46	.72	.85	.92	.10	.25	.49	.75	.90
GLS_u	.14	.38	.69	.91	.96	.10	.24	.49	.75	.92
REC	.18	.50	.82	.97	1.0	.10	.24	.49	.76	.92
WS	.19	.50	.83	.97	1.0	.10	.25	.50	.77	.93
MAX	.19	.50	.83	.97	1.0	.10	.24	.50	.76	.93
GLS^{WS}	.19	.47	.72	.86	.92	.10	.25	.49	.74	.90
GLS_u^{WS}	.19	.51	.83	.97	1.0	.10	.25	.50	.77	.93
REC^{WS}	.18	.50	.82	.97	1.0	.10	.25	.50	.76	.93
GLS^{MAX}	.19	.47	.72	.86	.92	.10	.25	.49	.75	.90
GLS_u^{MAX}	.18	.50	.82	.97	1.0	.10	.24	.49	.76	.93
REC^{MAX}	.19	.50	.82	.97	1.0	.10	.25	.50	.76	.93

Table 3. Finite sample power at nominal 0.05-level, $p_{\max} = 4$.

$T = 75$.

	constant case					trend case				
c	-5	-10	-15	-20	-25	-5	-10	-15	-20	-25
DF	.13	.30	.54	.74	.84	.08	.17	.32	.51	.69
GLS	.17	.40	.61	.72	.77	.09	.20	.38	.58	.73
GLS_u	.15	.34	.59	.78	.86	.09	.20	.38	.59	.74
REC	.18	.43	.69	.82	.88	.09	.21	.40	.60	.74
WS	.17	.40	.62	.76	.84	.09	.19	.36	.53	.67
MAX	.18	.43	.71	.86	.92	.09	.20	.40	.62	.79
GLS^{WS}	.17	.37	.54	.64	.67	.09	.20	.36	.51	.62
GLS_u^{WS}	.17	.40	.62	.76	.83	.09	.20	.36	.53	.65
REC^{WS}	.17	.40	.61	.75	.82	.09	.20	.35	.52	.63
GLS^{MAX}	.18	.42	.64	.75	.80	.09	.21	.41	.62	.77
GLS_u^{MAX}	.18	.44	.71	.85	.91	.09	.22	.42	.64	.78
REC^{MAX}	.18	.43	.71	.86	.91	.09	.21	.41	.63	.79

$T = 150$.

	constant case					trend case				
c	-5	-10	-15	-20	-25	-5	-10	-15	-20	-25
DF	.13	.32	.57	.78	.89	.09	.19	.36	.56	.74
GLS	.18	.41	.61	.83	.80	.09	.22	.41	.61	.75
GLS_u	.15	.37	.63	.82	.91	.10	.22	.41	.63	.79
REC	.18	.45	.72	.87	.93	.10	.22	.42	.64	.79
WS	.18	.44	.70	.85	.93	.09	.21	.40	.60	.75
MAX	.18	.46	.74	.90	.96	.10	.22	.43	.66	.83
GLS^{WS}	.18	.40	.59	.70	.77	.09	.22	.39	.56	.69
GLS_u^{WS}	.18	.44	.70	.85	.93	.09	.21	.40	.59	.73
REC^{WS}	.17	.43	.69	.84	.92	.09	.21	.39	.58	.72
GLS^{MAX}	.18	.42	.63	.76	.82	.10	.23	.43	.64	.78
GLS_u^{MAX}	.18	.45	.74	.89	.95	.10	.22	.44	.66	.81
REC^{MAX}	.18	.46	.74	.90	.96	.10	.22	.44	.65	.82

Table 4. Finite sample size at nominal 0.05-level, $p_{\max} = 4$.

$T = 75$.

	constant case				trend case			
ϕ, θ	0.5,0	-0.5,0	0,0.5	0,-0.5	0.5,0	-0.5,0	0,0.5	0,-0.5
DF	.04	.05	.12	.05	.04	.05	.19	.05
GLS	.04	.05	.11	.05	.04	.05	.18	.05
GLS_u	.05	.05	.13	.06	.04	.05	.19	.06
REC	.04	.05	.13	.05	.04	.05	.18	.05
WS	.05	.05	.11	.05	.04	.05	.16	.05
MAX	.04	.05	.13	.05	.04	.05	.18	.05
GLS^{WS}	.05	.04	.10	.05	.04	.04	.14	.04
GLS_u^{WS}	.05	.05	.11	.05	.05	.04	.15	.04
REC^{WS}	.04	.05	.11	.04	.04	.05	.15	.04
GLS^{MAX}	.04	.05	.11	.05	.04	.05	.16	.05
GLS_u^{MAX}	.04	.05	.13	.05	.03	.05	.17	.04
REC^{MAX}	.04	.04	.13	.05	.04	.05	.17	.05

$T = 150$.

	constant case				trend case			
ϕ, θ	0.5,0	-0.5,0	0,0.5	0,-0.5	0.5,0	-0.5,0	0,0.5	0,-0.5
DF	.05	.05	.09	.05	.05	.05	.12	.05
GLS	.05	.05	.09	.05	.05	.05	.12	.05
GLS_u	.05	.05	.10	.05	.05	.05	.12	.05
REC	.05	.05	.10	.04	.04	.05	.12	.05
WS	.05	.05	.09	.04	.05	.05	.11	.04
MAX	.05	.05	.10	.04	.04	.05	.12	.05
GLS^{WS}	.05	.05	.08	.04	.05	.05	.10	.04
GLS_u^{WS}	.05	.05	.09	.04	.05	.05	.11	.04
REC^{WS}	.04	.05	.09	.04	.04	.05	.11	.04
GLS^{MAX}	.05	.05	.09	.05	.04	.05	.11	.05
GLS_u^{MAX}	.04	.05	.10	.04	.04	.05	.11	.04
REC^{MAX}	.05	.05	.10	.04	.05	.05	.12	.05

Table 5. Finite sample size and power at nominal 0.05-level, constant case, $p = 0$, $T = 75$.

	$\chi^2(1) - 1$						$t(5)$					
c	0	-5	-10	-15	-20	-25	0	-5	-10	-15	-20	-25
DF	.05	.12	.34	.66	.89	.98	.05	.12	.34	.66	.89	.98
GLS	.04	.18	.51	.79	.92	.95	.05	.18	.50	.78	.91	.96
GLS_u	.05	.14	.40	.73	.93	.99	.05	.15	.40	.73	.93	.99
REC	.04	.18	.51	.84	.96	.99	.05	.18	.51	.84	.97	1.0
WS	.05	.18	.52	.84	.96	.99	.05	.18	.52	.85	.97	1.0
MAX	.05	.18	.52	.84	.96	.99	.05	.18	.51	.84	.97	1.0
GLS^{WS}	.04	.18	.51	.79	.91	.94	.05	.18	.50	.77	.90	.95
GLS_u^{WS}	.05	.18	.52	.84	.96	.99	.05	.18	.52	.84	.97	1.0
REC^{WS}	.04	.17	.51	.84	.96	.99	.05	.18	.51	.83	.97	1.0
GLS^{MAX}	.04	.18	.51	.80	.92	.95	.05	.18	.50	.78	.91	.96
GLS_u^{MAX}	.04	.18	.52	.84	.96	.99	.05	.18	.51	.84	.97	1.0
REC^{MAX}	.04	.17	.51	.84	.97	.99	.05	.18	.51	.84	.97	1.0