UNIVERSITY OF NOTTINGHAM

Discussion Papers in Economics

Discussion Paper
No. 03/09

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

by Stephen Leybourne, Tae-Hwan Kim and Paul Newbold

May 2003 DP 03/09

ISSN 1360-2438

UNIVERSITY OF NOTTINGHAM

Discussion Papers in Economics

Discussion Paper
No. 03/09

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

by Stephen Leybourne, Tae-Hwan Kim and Paul Newbold

Stephen Leybourne and Paul Newbold are Professors and Tae-Hwan Kim is Lecturer, all in the School of Economics, University of Nottingham

---

May 2003

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; 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 Fuller (1976) and 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 concentrates primarily on $\bar{c} = -10$, and we adopt this choice in our analysis of the tests.

3. Taylor (2000) and Shin and So (2001), motivated by So and Shin (1999), consider 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, explicitly or implicitly, OLS detrending, but exploit the fact that, under the alternative hypothesis of trend stationarity, any backward-looking ARMA model has, apart from the trend, an equivalent forward-looking representation. 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 minimisation 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 $v_t = y_{T+1-t}$ $(t = 1, ..., T)$. 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 $\max(DF_f, DF_r)$.

We note that the extension of all five of these tests to the "augmented" case, where lagged first differences are incorporated, is quite straightforward.

With the exception of the MAX test, the limiting null distributions of these five modified tests are all given in the cited literature. We fill this gap in Section 2 of the paper. 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 (2000), 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 so-called "local-to-unity asymptotics" approach. This limiting distribution is given for the MAX test in Section 3. Those for the other tests are given by, or can be directly inferred from, Elliott *et al* (1996), Elliott (1999), and Taylor (2000). We go on to calculate local asymptotic power for the five tests, comparing the results with the power envelope, given by Elliott (1999).

It emerges from the calculations of Section 3 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 is indistinguishable in reasonable sized samples. Accordingly, in Section 4 we report results of an extensive simulation study on the finite sample size and power 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 NULL DISTRIBUTIONS OF THE MAX TEST STATISTICS

Suppose that the time series y_t is generated through

$$y_t = y_{t-1} + \sum_{j=1}^{p-1} \phi_j \Delta y_{t-j} + \epsilon_t \tag{3}$$

where the roots of $1 - \sum_{j=1}^{p-1} \phi_j z^j = 0$ lie outside of the unit circle. Without loss of generality, the initial values y_t $(t = 0, -1, ..., -p+1)$ are assumed to exist. We impose the following assumption on ϵ_t.

Assumption 1. *ϵ_t is a martingale difference sequence and satisfies* $E(\epsilon_t^2|\epsilon_{t-1}, ...) = \sigma^2, E(|\epsilon_t|^i|\epsilon_{t-1}, ...) = \kappa_i$ $(i = 3, 4)$, *and* $\sup_t E(|\epsilon_t|^{4+\gamma}|\epsilon_{t-1}, ...) = \kappa < \infty$ *for some* $\gamma > 0$.

This assumption is standard in the unit root literature (see, for instance, Banerjee *et al*, 1992). The ADF regression based on the forward series y_t is

$$\Delta y_t = \hat{\gamma}' z_t + \hat{\rho} y_{t-1} + \sum_{j=1}^{p-1} \hat{\phi}_j \Delta y_{t-j} + \hat{\epsilon}_t \tag{4}$$

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

$$\Delta v_t = \tilde{\gamma}' z_t + \tilde{\rho} v_{t-1} + \sum_{j=1}^{p-1} \tilde{\phi}_j \Delta v_{t-j} + \tilde{\eta}_t. \tag{5}$$

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 null distribution of the MAX test statistics.

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) $z_t = 1$,

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

where

$$\begin{aligned} F_{n0} &= \frac{0.5\{W(1)^2 - 1\} - HW(1)}{(G - H^2)^{1/2}} \\ R_{n0} &= \frac{-0.5\{W(1)^2 + 1\} + HW(1)}{(G - H^2)^{1/2}} \\ H &= \int_0^1 W(r)dr \\ G &= \int_0^1 W(r)^2 dr. \end{aligned}$$

Here, $W(r)$ is a standard Brownian motion process.

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

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

where

$$\begin{aligned} F_{n1} &= \frac{0.5\{W(1)^2 - 1\} - 6MW(1) + 2HW(1) + 12HM - 6H^2}{(G - 12M^2 + 12HM - 4H^2)^{1/2}} \\ R_{n1} &= \frac{-0.5\{W(1)^2 + 1\} + 6MW(1) - 2HW(1) - 12HM + 6H^2}{(G - 12M^2 + 12HM - 4H^2)^{1/2}} \\ M &= \int_0^1 rW(r)dr. \end{aligned}$$

The two marginal limiting distributions for the forward and reverse $t-$statistics in each case are easily shown to be the same. It is straightforward to see that in the constant only case the limiting null distribution is invariant to starting values, and that in the linear trend case this distribution is invariant both to starting values and any drift in the random walk generating process.

3. ASYMPTOTIC DISTRIBUTIONS OF THE *MAX* STATISTIC UNDER THE LOCAL ALTERNATIVE

Here we consider a simplified data generating process given by

$$\begin{aligned} y_t &= \rho y_{t-1} + \epsilon_t \qquad t = 1, 2, ..., T \\ \rho &= 1 + \frac{c}{T} \end{aligned} \tag{6}$$

where $c \in (-\infty, 0)$.

Assumption 2. (i) y_0 is $N(0, \sigma^2(1-\rho^2)^{-1})$, (ii) ϵ_t is i.i.d.$(0, \sigma^2)$ and (iii) y_0 is uncorrelated with ϵ_t, $t \geq 1$.

Consider the following forward and reverse regression equations:

$$\begin{aligned} y_t &= \hat{\gamma}_0 + \hat{\rho} y_{t-1} + \hat{\epsilon}_t \\ v_t &= \tilde{\delta}_0 + \tilde{\rho} v_{t-1} + \tilde{\eta}_t \end{aligned} \tag{7}$$

where $v_t = y_{T+1-t}$. Let DF_{f0} and DF_{r0} be the $t-$statistics for testing $\rho = 1$ from the forward and the reversed regression equations in (7) respectively.

Theorem 2 *Suppose that y_t is generated by the DGP in (6) and Assumption 2 holds. Then,*

$$\begin{aligned} DF_{f0} &\Rightarrow \frac{0.5\{J_c(1)^2 - Z_c^2 - 1\} - H_c\{J_c(1) - Z_c\}}{(G_c - H_c^2)^{1/2}} = F_{a0} \\ DF_{r0} &\Rightarrow \frac{-0.5\{J_c(1)^2 - Z_c^2 + 1\} + H_c\{J_c(1) - Z_c\}}{(G_c - H_c^2)^{1/2}} = R_{a0} \\ MAX &= \max(DF_{f0}, DF_{r0}) \Rightarrow \max(F_{a0}, R_{a0}) \end{aligned}$$

where

$$\begin{aligned} Z_c &= N(0, (-2c)^{-1}) \\ J_c(r) &= W_c(r) + \exp(rc) Z_c \\ H_c &= \int_0^1 J_c(r) dr \\ G_c &= \int_0^1 J_c(r)^2 dr. \end{aligned}$$

Here, $W_c(r)$ is an Ornstein-Uhlenbeck process.

Consider the following forward and reverse regression equations:

$$\begin{aligned} y_t &= \hat{\gamma}_0 + \hat{\gamma}_1 t + \hat{\rho} y_{t-1} + \hat{\epsilon}_t \\ v_t &= \tilde{\delta}_0 + \tilde{\delta}_1 t + \tilde{\rho} v_{t-1} + \tilde{\eta}_t. \end{aligned} \tag{8}$$

Let DF_{f1} and DF_{r1} be the $t-$statistics for testing $\rho = 1$ from the forward and the reversed regression equations in (8) respectively.

Theorem 3 *Suppose that y_t is generated by the DGP in (6) and Assumption 2 holds. Then,*

$$\begin{aligned} DF_{f1} &\Rightarrow \frac{0.5\{J_c(1)^2 - Z_c^2 - 1\} - 6M_c\{J_c(1) + Z_c\} + 2H_c\{J_c(1) + 2Z_c\} + 12H_cM_c - 6H_c^2}{(G_c - 12M_c^2 + 12H_cM_c - 4H_c^2)^{1/2}} \\ &= F_{a1} \\ DF_{r1} &\Rightarrow \frac{-0.5\{J_c(1)^2 - Z_c^2 + 1\} + 6M_c\{J_c(1) + Z_c\} - 2H_c\{J_c(1) + 2Z_c\} - 12H_cM_c + 6H_c^2}{(G_c - 12M_c^2 + 12H_cM_c - 4H_c^2)^{1/2}} \\ &= R_{a1} \\ MAX &= \max(DF_{f1}, DF_{r1}) \Rightarrow \max(F_{a1}, R_{a1}) \end{aligned}$$

where

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

Again, the two marginal limiting distributions in both the constant only case and the linear trend case are easily shown to be the same, the limiting distribution in Theorem 2 is invariant to any mean and that in Theorem 3 is invariant to any linear trend in the true generating process. 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. Some results for 5%-level tests are shown in Tables 1 and 2, where ENV denotes the asymptotic Gaussian power envelope, taken from Elliott (1999). These results were obtained by simulating 50,000 replications of the appropriate limiting functionals, using series of 5,000 Gaussian white noise innovations. Here and throughout, all calculations were programmed in GAUSS. Notice first that, with the exception of the constant case for $c = -20$, 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. The gains for the modified tests tend to be greater in the constant case than in the linear trend case. It emerges that the REC, WS, and

MAX tests have asymptotic local power that is always as high as, and sometimes much higher than the two tests based on generalised least squares detrending. Moreover, the performance of the three superior tests is very close to the power envelope, suggesting that in very large samples, when the innovations are Gaussian, it is not possible to improve noticeably on the powers of these tests.

Of course, asymptotic local power calculations may provide an imperfect guide to what will be found for sample sizes of practical interest. Accordingly, in the next section we report simulation results on finite sample power, as well as on the reliability of test sizes in circumstances such as the augmentation of fitted models through the incorporation of lagged changes and non-normal innovations.

4. SMALL SAMPLE SIMULATIONS

The augmented version of the Dickey-Fuller test is based on fitting the model

$$y_t = \gamma' z_t + \rho y_{t-1} + \sum_{j=1}^{p-1} \phi_j \Delta y_{t-j} + \epsilon_t \tag{9}$$

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 in tests of reliable size 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. However, published results have often made it difficult to form a clear picture, as considerations of power can be obscured by considerations of size reliability. In the usual application of (9), critical values for tests based on samples of T observations are derived from the simulation of random walks, with Gaussian innovations, of that number of observations, with the number of lagged first differences in the fitted model fixed at $p = 1$. Test sizes are then strictly correct only for the case of testing a random walk with Gaussian innovations against the alternative of a stationary first order autoregression. Actual test sizes depend not only on T, but also on p and the specific values of the true ϕ_j, as well as the distribution of ϵ_t. This last consideration has sometimes motivated the use of asymptotic critical values, on the grounds that these will be correct for very large samples whatever the innovation distribution. The situation is further complicated by the fact that, as a practical matter, p will be unknown and an appropriate value must be selected on the basis of data evidence. Indeed, the true generating process may include moving average terms, in which case any chosen model of the form (9) will necessarily be an approximation to that process. Simulation of a variety of generating processes allied with test applications in which p is selected through a data-dependent rule undoubtedly reflects practical reality. Unfortunately, the results of such exercises confound power with size reliability, whatever critical values are used. We note that size-adjusted power, while of some theoretical interest, has limited practical relevance since it generally cannot be achieved in practice. Since none of the modified tests is motivated by size reliability considerations, we feel that it is important to compare their small sample power in a case where size is known to be correct.

This view leads us to emphasise the basic case in which the generating process is a random walk with Gaussian innovations under the null, and a stationary first order autoregression under the alternative, with tests based on the fitting of (9), setting $p = 1$. Subsequently we shall explore size reliability in some more elaborate cases. We analysed the DF test, its five modifications, and the six hybrid tests based on those modifications. Critical values for given sample sizes were determined through 20,000 replications. Thus the tests are known to be correctly sized for this basic case, so that the proportions of rejections under the alternative do indeed measure test power. Tables 3 and 4 show for 5%-level tests the outcomes for samples of 50, 100 and 200 observations - the former table for the constant only case, and the latter for the linear trend case. Here and throughout results on rejection proportions are based on 20,000 replications, the generating process under the alternative being truly stationary in the sense that the deviation from trend of the initial observation has variance $(1-\rho^2)^{-1}$, in the first order autoregressive case, with ϵ_t generated as standard normal.

The overall pattern of Tables 3 and 4 is quite similar, though as might be anticipated from the large sample results of the previous section, power differences among the tests are more pronounced in the constant only case than when a linear trend is incorporated. The usual Dickey-Fuller test is generally comfortably outperformed by all its competitors. Of those competitors, GLS_u is usually least powerful, though it is somewhat more powerful than GLS in the large sample/high power case, where both are inferior to their competitors. The relatively poor performance of GLS_u is somewhat ironic, as it was designed with the strictly stationary alternative analysed here in mind. There is very little to choose among the REC, WS, and MAX tests, with WS generally marginally the most powerful. The six hybrid tests have no noticeable advantage over these three, though the hybrids of GLS_u perform relatively well where GLS_u itself does not. The same is not true however of the hybrids of GLS.

Although the results of Tables 3 and 4 suggest a very definite preference for some of these unit root tests over the others in the basic random walk versus stationary first order autoregression case, it is important to check also for size robustness in more elaborate cases. We begin with a generating model in which first differences follow a stationary first order autoregression with parameter ϕ, the innovations are Gaussian, and tests are based on fitted models of the form (9) with, correctly, $p = 2$. Here and elsewhere, in computing the MAX statistic, the same value of p is used in the forward and reverse regressions. The augmented variant of the WS test is specified in Pantula *et al* (1994). Standard finite sample critical values are used - that is, the same critical values as for the simulations of Tables 3 and 4. Tables 5 and 6 show empirical sizes for nominal 5% and 10%-level tests for samples of 50 and 100 observations. The former table is for the constant only case, and the latter for the linear trend case. With notable exceptions, empirical sizes are close to nominal sizes. It is particularly noteworthy that the WS and MAX tests are less over-sized than the usual Dickey-Fuller test in the extreme case $\phi = 0.9$. An interesting case of large discrepancies between empirical and nominal sizes occurs for the REC test and its hybrids. These can be quite seriously *under-sized* for positive ϕ. This suggests the possibility that these tests could generate relatively low rejection probabilities for particular stationary second order autoregressions.

We checked this conjecture by generating series of 50 and 100 observations, again with Gaussian innovations, from the stationary processes

$$(1 - \rho L)^2 y_t = \epsilon_t$$

where L is the lag operator. Tables 7 and 8 show proportions of rejections of the unit-root null hypothesis from nominal 5%-level tests, based on the same critical values as Tables 3-6. As predicted, REC and its hybrids are noticeably less likely to reject the null hypothesis than either WS or MAX, suggesting a preference for the latter pair. Properly speaking, the entries in these tables are not powers, as the test sizes are not precisely .05. Nevertheless, they do estimate what an analyst would obtain in practice, and in this sense are more informative than size-adjusted power. Moreover, as we saw in Tables 5 and 6, the WS and MAX tests are not seriously over-sized. To the contrary, it is the low empirical sizes of the REC test and its hybrids that is responsible for the apparently relatively poor performance of these test in Tables 7 and 8. Thus, if we were to base a choice of test on the power results of Tables 3,4,7 and 8, the WS and MAX tests and the two hybrids of GLS_u stand out. These four tests performed about equally well in these simulation experiments. In particular, the hybrids are not noticeably superior to WS and MAX alone, so that there seems to be little point in incurring further computational elaboration by employing them. Indeed, our results provide no evidence that OLS detrending is the culprit for the acknowledgedly suboptimal performance of the usual Dickey-Fuller tests, since WS explicitly and MAX implicitly involve such detrending, and neither is improved by alternative detrending methods.

We further explored robustness to lag order by basing tests on estimated models of the form (9) with $p = 6$ fitted to random walks with Gaussian white noise innovations. Empirical sizes for nominal 5%-level tests based on samples of 50 and 100 observations are given in Table 9. Of our two preferred tests, WS is approximately correctly sized while MAX is moderately under-sized. Also, as might perhaps be expected, given the results of Tables 5 and 6, REC and its hybrids are seriously under-sized.

Because finite sample critical values are strictly valid only under normality, while asymptotic critical values are valid asymptotically for very general innovation distributions, some analysts prefer to use the latter whatever the sample size. To explore possible consequences of doing so we generated random walks, with Gaussian innovations, of 50 and 100 observations, basing tests on the fitting of (9) with $p = 1$, but now employing the asymptotic critical values. Table 10 shows empirical sizes. Our two preferred tests are somewhat over-sized, WS more seriously so than MAX, by this criterion in the linear trend case. A more striking finding is that the GLS test and its two hybrids can be quite seriously over-sized, particularly in the constant only case, when asymptotic critical values are used. Such an approach then is susceptible to spurious rejections of the unit root null hypothesis.

Finally, to check robustness to non-normality, we simulated random walks of 50 observations using innovations from a highly skewed distribution, $\chi^2(1) - 1$, and also from a heavy-tailed distribution, Students' t with five degrees of freedom. Tests were based on fitting (9) with $p = 1$, using finite sample critical values derived from a normality assumption. Table 11 shows empirical sizes of nominal 5% and 10%-level tests in the constant only case. (Results for the linear trend case are very similar.) There is little difference among the tests - all tend to be mildly under-sized for both innovation distributions.

5. 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 forward and backward representations of covariance-stationary processes. 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 large sample and small sample properties of the tests. On the basis of our results, we recommend two tests - the WS test of Pantula *et al* (1994) and the MAX test of Leybourne (1995). Our power and size simulations suggest there is little to choose between the two, and that if one or the other is employed, there is no particular advantage in alternatives to OLS detrending. Both tests are quite easily programmed, the MAX being rather more straightforward, particularly in the general case where the Dickey-Fuller regression is augmented by lagged first differences. It also generalises more easily to the modification of Perron tests in the case of an additive outlier-type structural break.

APPENDIX

PROOF OF THEOREM 1. The limiting distributions of the forward regression statistics, F_{n0} and F_{n1}, are well known in the literature. They are functions of a standard Brownian motion process $W(r)$ defined as $\sigma^{-1}T^{-1/2}\sum_{t=1}^{rT}\epsilon_t \Rightarrow W(r)$. First, we consider the constant only case: $z_t = 1$. In this case, we have from the literature:

$$DF_f \Rightarrow F_{n0} = \frac{0.5\{W(1)^2-1\} - W(1)\int_0^1 W(r)dr}{\left[\int_0^1 W(r)^2 dr - \left\{\int_0^1 W(r)dr\right\}^2\right]^{1/2}}. \tag{10}$$

The first step to derive the limiting distribution of DF_r is to find the corresponding DGP for the reverse series v_t. Let $\phi(z) = 1-\sum_{j=1}^{p-1}\phi_j z^j$. Recalling that the DGP of Δy_t under the unit root hypothesis is

$$\begin{aligned}\Delta y_t &= \sum_{j=1}^{p-1}\phi_j \Delta y_{t-j} + \epsilon_t \\ &= \phi(L)^{-1}\epsilon_t\end{aligned} \tag{11}$$

we need to find a solution of η_t that satisfies the following equation:

$$\Delta\tilde{y}_t = \sum_{j=1}^{p-1}\phi_j \Delta\tilde{y}_{t-j} + \eta_t. \tag{12}$$

Noting that $\Delta v_{t-j} = -\Delta y_{T-t+2+j}$, (12) can be written as $\Delta y_t = \sum_{j=1}^{p-1}\phi_j \Delta y_{t+j}$ $-\eta_{T-t+2}$ which implies that $\Delta y_t = -\phi(L^{-1})^{-1}\eta_{T-t+2}$. Combining the last result with (11), the solution for η_t is given by

$$\eta_t = -\psi(L)\varepsilon_{T-t+2} = \sum_{j=-(p-1)}^{\infty}\psi_j \varepsilon_{T-t+2-j} \tag{13}$$

where $\psi(L) = \phi(L^{-1})\phi(L)^{-1}$. It can be easily shown by the stationarity condition on $\phi(L)$ that $\sum_{j=-(p-1)}^{\infty} j|\psi_j| < \infty$. Let $N(s)$ be the Brownian motion defined by the partial sum of η_t (its existence will be shown below): that is $\sigma^{-1}T^{-1/2}\sum_{t=1}^{rT}\eta_t \Rightarrow N(r)$. We can make the following observations: (i) the only difference between the two DGPs in (11) and (12) is the error terms (ϵ_t and η_t) and (ii) the regressions in (4) and (5) are of the identical form. Hence, as the forward statistic DF_f is a function of $\sigma^{-1}T^{-1/2}\sum_{t=1}^{T}\epsilon_t$ and $\sigma^{-2}T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_t^2$ denoted by $h(\cdot,\cdot)$, the reverse statistic DF_r should have the same function $h(\cdot,\cdot)$, but the arguments are now $\sigma^{-1}T^{-1/2}\sum_{t=1}^{T}\eta_t$ and $\sigma^{-2}T^{-1}\sum_{t=1}^{T}\tilde{\eta}_t^2$ instead. Using the fact that the function $h(\cdot,\cdot)$ is continuous in both arguments and the two residual variance estimators, $T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_t^2$ and $T^{-1}\sum_{t=1}^{T}\tilde{\eta}_t^2$, converge to σ^2 in probability, we have

$$\begin{aligned} DF_f &= h\left(\sigma^{-1}T^{-1/2}\sum_{t=1}^{T}\varepsilon_t, \sigma^{-2}T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_t^2\right) \Rightarrow h(W(1),1), \\ DF_r &= h\left(\sigma^{-1}T^{-1/2}\sum_{t=1}^{T}\eta_t, \sigma^{-2}T^{-1}\sum_{t=1}^{T}\tilde{\epsilon}_t^2\right) \Rightarrow h(N(1),1). \end{aligned}$$

Since the expression of $h(W(1))$ is in (10), the limit of the reverse statistic $h(N(1))$ is now obtained by replacing $W(1)$ with $N(1)$ in (10):

$$h(N(1)) = R_{n0} = \frac{0.5(N(1)^2-1) - N(1)\int_0^1 N(r)dr}{\left[\int_0^1 N(r)^2dr - \left\{\int_0^1 N(r)dr\right\}^2\right]^{1/2}}. \tag{14}$$

Comparing (10) and (14), it can be easily concluded that the two limiting distributions for the forward and reverse statistics are the same marginally. Now we need to prove the existence of the new Brownian motion $N(\cdot)$. Using the results in (13) and Beveridge and Nelson (1981), it is easily seen that

$$\begin{aligned} \sigma^{-1}T^{-1/2}\sum_{t=1}^{rT}\eta_t &= -\sigma^{-1}T^{-1/2}\psi(1)\sum_{t=1}^{rT}\varepsilon_{T-t+2} + o_p(1) \qquad (15) \\ &= -\left\{\sigma^{-1}T^{-1/2}\psi(1)\sum_{t=1}^{T}\varepsilon_t - \sigma^{-1}T^{-1/2}\psi(1)\sum_{t=1}^{(1-r)T}\varepsilon_t\right\} + o_p(1) \\ &\Rightarrow -\{W(1) - W(1-r)\}. \end{aligned}$$

The dependence of η_t on ϕ_j disappears in the limit since $\psi(1) = 1$. Hence, the new Brownian motion $N(r)$ is in fact given by $-\{W(1) - W(1-r)\}$. Therefore, in order to express the limit of the reverse statistic DF_r in terms of the original Brownian motion $W(\cdot)$, we simply replace $N(r)$ with $-\{W(1) - W(1-r)\}$. More precisely, we have

$$\begin{aligned} N(1) &= -W(1), \\ \int_0^1 N(r)dr &= -W(1) + \int_0^1 W(r)dr, \\ \int_0^1 N(r)^2dr &= W(1)^2 - 2W(1)\int_0^1 W(r)dr + \int_0^1 W(r)^2dr. \end{aligned}$$

Substituting these results into (14) and collecting terms gives that

$$\frac{-0.5(W(1)^2+1)+W(1)\int_0^1 W(r)dr}{\left[\int_0^1 W(r)^2dr-\left\{\int_0^1 W(r)dr\right\}^2\right]^{1/2}}$$

which is the expression given for R_{n0} in Theorem 1. Finally, the continuous mapping theorem delivers the required result: $MAX \Rightarrow \max(F_{n0}, R_{n0})$.

The proof for the linear trend case is entirely analogous. The only additional result required is:

$$\int_0^1 rN(r)dr = -0.5W(1) + \int_0^1 W(r)dr - \int_0^1 rW(r)dr.$$

Using this additional result, the same argument is applied to R_{n1} and delivers $MAX \Rightarrow \max(F_{n1}, R_{n1})$.

PROOF OF THEOREM 2. First note that $y_{rT} = \sum_{s=1}^{rT}\rho^{rT-s}\epsilon_s + \rho^{rT}y_0$. Since $T^{-1/2}\sum_{s=1}^{rT}\rho^{rT-s}\epsilon_s \Rightarrow \sigma W_c(r)$ [by Phillips (1987)], $\rho^{rT} \to \exp(rc)$, and $T^{-1/2}y_0 \Rightarrow \sigma Z_c$ [by Elliott (1999)], we have $T^{-1/2}y_{rT} \Rightarrow \sigma W_c(r) + \sigma\exp(rc)Z_c = \sigma J_c(r)$. By the continuous mapping theorem, we can immediately obtain $T^{-3/2}\sum_2^T y_{t-1} \Rightarrow \sigma\int_0^1 J_c(r)dr = \sigma H_c$ and $T^{-2}\sum_2^T y_{t-1}^2 \Rightarrow \sigma^2\int_0^1 J_c^2(r)dr = \sigma^2 G_c$. Next consider

$$\begin{aligned} T^{-1}\sum_2^T y_{t-1}\epsilon_t &= (2\rho)^{-1}(T^{-1}y_T^2 - T^{-1}y_1^2 - T^{-1}\sum_2^T \epsilon_t^2 - 2cT^{-2}\sum_2^T y_{t-1}^2) + o_p(1) \\ &\Rightarrow \sigma^2\{0.5(J_c(1)^2 - Z_c^2 - 1) - cG_c\} = \sigma^2 E_c. \end{aligned}$$

Let $\hat{\beta} = (\hat{\gamma}_0, \hat{\rho})'$ from the forward regression in (7), $\beta = (0,\rho)'$ and $D_T = diag(T^{1/2}, T)$. Then we have

$$\hat{\beta}-\beta = \begin{bmatrix} T-1 & \sum_2^T y_{t-1} \\ \sum_2^T y_{t-1} & \sum_2^T y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_2^T \epsilon_t \\ \sum_2^T y_{t-1}\epsilon_t \end{bmatrix}$$

which implies that $D_T(\hat{\beta}-\beta) \Rightarrow A_f^{-1}B_f$ 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 $T^{-1}\sum_2^T \hat{\epsilon}_t^2 \to_p \sigma^2$, 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 $DF_{f0} = c\{Tv\hat{a}r(\hat{\rho})\}^{-1/2} + (\hat{\rho}-\rho)v\hat{a}r(\hat{\rho})^{-1/2} \Rightarrow c(G_c - H_c^2)^{1/2} + \{E_c - H_cW(1)\}(G_c - H_c^2)^{-1/2}$ which simplifies to the result in the theorem.

Now we consider the reverse regression. Let $\tilde{\beta} = (\tilde{\delta}_0, \tilde{\rho})'$ from the reverse regression in (7). 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

$$\tilde{\beta} - \beta = \begin{bmatrix} T-1 & \sum_2^T v_{t-1} \\ \sum_2^T v_{t-1} & \sum_2^T v_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_2^T \eta_t \\ \sum_2^T v_{t-1}\eta_t \end{bmatrix}.$$

It can be shown that $T^{-3/2}\sum_2^T v_{t-1} \Rightarrow \sigma H_c, T^{-2}\sum_2^T v_{t-1}^2 \Rightarrow \sigma^2 G_c, T^{-1/2}\sum_2^T \eta_t \Rightarrow -\sigma(W(1) + 2cH_c)$ and $T^{-1}\sum_2^T v_{t-1}\eta_t \Rightarrow -\sigma^2(E_c + 1 + 2cG_c)$. Hence, we have $D_T(\tilde{\beta} - \beta) \Rightarrow A_r^{-1} B_r$ where

$$A_r = A_f, \quad B_r = \begin{bmatrix} -\sigma\{W(1) + 2cH_c\} \\ -\sigma^2(E_c + 1 + 2cG_c) \end{bmatrix}. \tag{16}$$

The above result together with $T^{-1}\sum_2^T \tilde{\eta}_t^2 \rightarrow_p \sigma^2$ 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 $DF_{r0} = c\{Tv\hat{a}r(\tilde{\rho})\}^{-1/2} + (\tilde{\rho}-\rho)v\hat{a}r(\tilde{\rho})^{-1/2} \Rightarrow c(G_c - H_c^2)^{1/2} + \{-E_c - 1 - 2cG_c + H_c(W(1) + 2cH_c)\}(G_c - H_c^2)^{-1/2}$ which simplifies to the result in the theorem. Once we obtain these two results, then we have $\max(DF_{f0}, DF_{r0}) \Rightarrow \max(F_{a0}, R_{a0})$ by the continuous mapping theorem.

PROOF OF THEOREM 3. Since the proof is very similar to that of Theorem 2, its detail is not presented. The only additional part is to establish the following limits:

$$\begin{aligned}
T^{-5/2}\sum_2^T t y_{t-1} &\Rightarrow \sigma\int_0^1 rJ_c(r)dr = \sigma M_c \\
T^{-3/2}\sum_2^T t\epsilon_t &= T^{-1/2}y_T - T^{-3/2}\sum_2^T y_{t-1} - cT^{-5/2}\sum_2^T t y_{t-1} + o_p(1) \\
&\Rightarrow \sigma\{J_c(1) - H_c - cM_c\} \\
T^{-5/2}\sum_2^T t v_{t-1} &= T^{-3/2}\sum_2^T y_{t-1} - T^{-5/2}\sum_2^T t y_{t-1} + o_p(1) \\
&\Rightarrow -\sigma(M_c - H_c) \\
T^{-3/2}\sum_2^T t\eta_t &= -2cT^{-3/2}\sum_1^{T-1} y_t + 2cT^{-5/2}\sum_1^{T-1} t y_t - \rho T^{-1/2}\sum_2^T \epsilon_{t-1} \\
&\quad + \rho T^{-3/2}\sum_2^T t\epsilon_{t-1} + o_p(1) \\
&\Rightarrow \sigma\{J_c(1) - (1+2c)H_c + cM_c - W(1)\}.
\end{aligned}$$

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] BEVERIDGE, S. and NELSON, C.R. (1981) A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle.' *Journal of Monetary Economics* 7, 151-174.

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

[4] 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.

[5] 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.

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

[7] FULLER, W.A., 1976. Introduction to Statistical Time Series. Wiley, New York.

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

[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] SO, B.S. and SHIN, D.W. (1999) Recursive mean adjustment in time series inferences. *Statistics and Probability Letters* 43, 65-73.

[14] TAYLOR, A.M.R. (2000) Regression-based unit root tests with recursive mean adjustment for seasonal and non-seasonal time series. Department of Economics Discussion Paper, University of Birmingham.

Table 1. Asymptotic local power at nominal 0.05-level, constant only case.

c	-5	-10	-15	-20
ENV	.20	.52	.83	.97
DF	.13	.33	.62	.86
GLS	.20	.44	.64	.77
GLS_u	.15	.39	.70	.90
REC	.19	.50	.81	.96
WS	.20	.52	.83	.96
MAX	.20	.50	.82	.96

Table 2. Asymptotic local power at nominal 0.05-level, linear trend case.

c	-5	-10	-15	-20
ENV	.10	.24	.49	.74
DF	.08	.19	.38	.62
GLS	.10	.24	.46	.68
GLS_u	.10	.24	.48	.74
REC	.10	.24	.49	.74
WS	.10	.24	.49	.74
MAX	.10	.24	.49	.74

Table 3. Power of tests at nominal 0.05-level, constant only, $p = 1$, DGP: $(1 - \rho L)y_t = \epsilon_t$.

(a) $T = 50$.

ρ	.950	.900	.850	.800	.750	.700	.650	.600
DF	.08	.13	.22	.35	.51	.67	.80	.90
GLS	.11	.20	.35	.53	.71	.85	.94	.98
GLS_u	.09	.16	.26	.41	.58	.73	.86	.93
REC	.10	.20	.34	.52	.70	.84	.93	.97
WS	.11	.20	.35	.54	.72	.85	.94	.98
MAX	.10	.20	.34	.52	.70	.84	.93	.97
$GLS - WS$	.11	.20	.35	.53	.71	.84	.92	.97
$GLS_u - WS$	.11	.20	.35	.54	.72	.86	.94	.98
$REC - WS$	.10	.20	.34	.52	.70	.84	.93	.97
$GLS - MAX$	.11	.20	.35	.54	.72	.86	.94	.98
$GLS_u - MAX$	.10	.20	.34	.52	.70	.85	.93	.97
$REC - MAX$	.11	.20	.35	.53	.71	.85	.94	.97

(b) $T = 100$.

ρ	.975	.950	.925	.900	.875	.850	.825	.800
DF	.08	.13	.22	.34	.49	.65	.78	.88
GLS	.10	.19	.33	.51	.67	.81	.90	.95
GLS_u	.08	.15	.25	.39	.56	.72	.84	.92
REC	.10	.19	.33	.51	.69	.83	.92	.97
WS	.10	.19	.34	.52	.70	.84	.93	.97
MAX	.10	.19	.33	.51	.69	.83	.92	.97
$GLS - WS$	.10	.19	.33	.50	.66	.79	.88	.93
$GLS_u - WS$	.10	.19	.34	.52	.70	.84	.93	.97
$REC - WS$	.10	.19	.33	.51	.69	.83	.92	.97
$GLS - MAX$	.10	.19	.33	.51	.67	.80	.89	.94
$GLS_u - MAX$	.10	.19	.33	.51	.69	.83	.92	.97
$REC - MAX$	.10	.19	.33	.52	.70	.84	.93	.97

(c) $T = 200$.

ρ	.975	.950	.925	.900	.875
DF	.13	.34	.64	.87	.97
GLS	.20	.50	.76	.89	.95
GLS_u	.15	.40	.71	.92	.99
REC	.19	.51	.82	.97	1.0
WS	.20	.52	.84	.97	1.0
MAX	.20	.51	.83	.97	1.0
$GLS-WS$	.20	.50	.75	.88	.95
GLS_u-WS	.20	.53	.84	.97	1.0
$REC-WS$	.19	.51	.82	.97	1.0
$GLS-MAX$	.20	.49	.74	.88	.95
GLS_u-MAX	.19	.51	.83	.97	1.0
$REC-MAX$	.20	.52	.84	.97	1.0

Table 4. Power of tests at nominal 0.05-level, linear trend, $p = 1$,
DGP: $(1 - \rho L)y_t = \epsilon_t$.

(a) $T = 50$.

ρ	.850	.800	.750	.700	.650	.600	.550	.500
DF	.13	.19	.29	.41	.55	.67	.79	.88
GLS	.16	.25	.37	.52	.66	.78	.88	.94
GLS_u	.15	.24	.36	.50	.65	.77	.87	.94
REC	.16	.25	.38	.52	.67	.79	.88	.94
WS	.16	.25	.38	.53	.68	.80	.89	.95
MAX	.16	.25	.38	.53	.67	.79	.88	.94
$GLS - WS$	.16	.25	.38	.52	.66	.78	.87	.93
$GLS_u - WS$	.16	.25	.38	.53	.68	.79	.89	.95
$REC - WS$	.16	.25	.38	.52	.67	.79	.88	.94
$GLS - MAX$	.16	.25	.38	.53	.67	.78	.88	.93
$GLS_u - MAX$	.16	.25	.37	.53	.67	.79	.89	.94
$REC - MAX$	.16	.25	.38	.53	.68	.79	.89	.95

(b) $T = 100$.

ρ	.925	.900	.875	.850	.825	.800	.775	.750
DF	.13	.20	.29	.40	.53	.66	.77	.86
GLS	.16	.25	.37	.51	.64	.75	.84	.90
GLS_u	.16	.25	.36	.50	.64	.77	.86	.93
REC	.16	.25	.37	.51	.65	.78	.87	.93
WS	.16	.25	.38	.52	.66	.78	.87	.94
MAX	.16	.25	.37	.51	.65	.78	.88	.93
$GLS - WS$	.16	.25	.37	.50	.63	.75	.84	.89
$GLS_u - WS$	.16	.25	.38	.52	.66	.78	.88	.94
$REC - WS$	.16	.25	.37	.51	.65	.77	.87	.93
$GLS - MAX$	.16	.25	.37	.50	.63	.74	.83	.89
$GLS_u - MAX$	.16	.25	.37	.51	.66	.78	.88	.94
$REC - MAX$	.16	.25	.37	.51	.66	.78	.87	.94

(c) $T = 200$.

ρ	.975	.950	.925	.900	.875	.850
DF	.09	.20	.39	.64	.85	.96
GLS	.10	.25	.48	.72	.87	.94
GLS_u	.10	.24	.48	.75	.91	.98
REC	.10	.25	.49	.75	.92	.98
WS	.10	.25	.51	.76	.92	.99
MAX	.10	.25	.49	.76	.92	.98
$GLS - WS$	.10	.25	.48	.72	.87	.94
$GLS_u - WS$	.10	.25	.51	.76	.93	.99
$REC - WS$	.10	.25	.49	.75	.92	.98
$GLS - MAX$	.10	.25	.48	.70	.86	.94
$GLS_u - MAX$	.10	.25	.50	.76	.93	.99
$REC - MAX$	.10	.25	.51	.76	.92	.98

Table 5. Size of tests at nominal 0.05 / 0.l0-level, constant only, $p = 2$,
DGP: $(1 - \phi L)(1 - L)y_t = \epsilon_t$.
(a) $T = 50$.

ϕ	-0.90	-0.50	0.0	0.50	0.90
DF	.048	.050	.050	.053	.071
	.096	.096	.099	.103	.133
GLS	.049	.049	.048	.050	.059
	.098	.099	.100	.100	.117
GLS_u	.045	.047	.046	.047	.049
	.090	.092	.093	.091	.098
REC	.056	.048	.045	.039	.027
	.112	.098	.093	.084	.059
WS	.052	.052	.051	.053	.056
	.102	.105	.104	.103	.106
MAX	.049	.050	.048	.049	.054
	.099	.101	.102	.102	.105
$GLS - WS$	.036	.048	.049	.052	.059
	.077	.099	.102	.104	.112
$GLS_u - WS$	.048	.051	.051	.053	.057
	.096	.104	.104	.104	.110
$REC - WS$	.043	.047	.045	.039	.026
	.089	.096	.093	.085	.059
$GLS - MAX$	.026	.042	.046	.050	.057
	.062	.090	.098	.102	.112
$GLS_u - MAX$	.045	.050	.049	.050	.057
	.095	.101	.103	.104	.113
$REC - MAX$	.057	.051	.045	.037	.018
	.112	.102	.095	.081	.044

(b) $T = 100$.

ϕ	-0.90	-0.50	0.0	0.50	0.90
DF	.049	.050	.049	.050	.060
	.099	.099	.097	.100	.116
GLS	.048	.050	.048	.048	.049
	.097	.100	.100	.101	.103
GLS_u	.047	.048	.048	.045	.046
	.095	.096	.094	.093	.093
REC	.055	.049	.045	.041	.028
	.106	.098	.094	.083	.061
WS	.051	.051	.049	.050	.050
	.102	.101	.101	.102	.100
MAX	.051	.052	.049	.049	.050
	.102	.102	.102	.101	.101
$GLS - WS$	.036	.048	.048	.048	.049
	.079	.098	.100	.102	.102
$GLS_u - WS$	.048	.051	.049	.050	.050
	.098	.100	.100	.100	.100
$REC - WS$	.047	.049	.045	.041	.028
	.095	.097	.094	.084	.061
$GLS - MAX$	.029	.045	.046	.047	.047
	.065	.088	.095	.099	.097
$GLS_u - MAX$	.048	.050	.050	.049	.051
	.099	.100	.101	.101	.103
$REC - MAX$	.054	.050	.046	.040	.022
	.107	.100	.097	.086	.052

Note: In each cell, upper entry is for nominal .05-level test and lower entry is for nominal .10-level test.

Table 6. Size of tests at nominal 0.05 / 0.10-level, linear trend, $p = 2$, DGP: $(1 - \phi L)(1 - L)y_t = \epsilon_t$.

(a) $T = 50$.

ϕ	-0.90	-0.50	0.0	0.50	0.90
DF	.049	.050	.051	.055	.065
	.097	.099	.103	.108	.127
GLS	.038	.047	.050	.051	.059
	.082	.096	.100	.102	.117
GLS_u	.036	.044	.046	.049	.055
	.074	.095	.096	.099	.113
REC	.047	.045	.042	.035	.017
	.098	.092	.084	.071	.037
WS	.051	.053	.054	.054	.053
	.104	.106	.105	.105	.103
MAX	.049	.049	.051	.050	.050
	.098	.101	.101	.099	.098
$GLS - WS$	.022	.043	.050	.053	.059
	.050	.090	.100	.104	.117
$GLS_u - WS$	.029	.047	.051	.053	.059
	.065	.097	.101	.105	.113
$REC - WS$	.049	.047	.044	.036	.016
	.102	.095	.087	.073	.036
$GLS - MAX$	.022	.039	.045	.049	.057
	.049	.082	.093	.100	.111
$GLS_u - MAX$	.029	.044	.049	.051	.058
	.064	.094	.098	.100	.114
$REC - MAX$	.048	.047	.042	.029	.008
	.101	.096	.083	.058	.019

(b) $T = 100$.

ϕ	-0.90	-0.50	0.0	0.50	0.90
DF	.050	.050	.051	.053	.060
	.098	.099	.103	.106	.120
GLS	.043	.049	.052	.053	.053
	.091	.098	.099	.101	.105
GLS_u	.041	.047	.050	.052	.055
	.086	.098	.102	.104	.110
REC	.049	.047	.045	.037	.018
	.101	.096	.091	.079	.040
WS	.050	.051	.054	.053	.050
	.102	.102	.104	.106	.101
MAX	.049	.049	.052	.052	.049
	.098	.099	.101	.102	.099
$GLS - WS$	.027	.045	.051	.053	.053
	.062	.092	.099	.103	.106
$GLS_u - WS$	.036	.048	.053	.053	.052
	.080	.098	.101	.105	.107
$REC - WS$	.050	.048	.046	.038	.018
	.102	.098	.093	.080	.039
$GLS - MAX$	.029	.045	.051	.053	.051
	.064	.092	.099	.101	.102
$GLS_u - MAX$	.036	.047	.052	.053	.053
	.079	.097	.100	.104	.108
$REC - MAX$	.050	.047	.046	.034	.011
	.101	.096	.090	.073	.026

See note to Table 5.

Table 7. Power of tests at nominal 0.05-level, constant only, $p = 2$, DGP: $(1 - \rho L)^2 y_t = \epsilon_t$.

	$T = 50$				$T = 100$			
ρ	.90	.80	.70	.60	.90	.80	.70	.60
DF	.09	.15	.26	.41	.14	.37	.69	.91
GLS	.11	.22	.39	.59	.21	.52	.82	.95
GLS_u	.09	.17	.30	.47	.16	.43	.75	.94
REC	.07	.16	.33	.54	.14	.47	.82	.97
WS	.12	.23	.40	.60	.21	.54	.86	.98
MAX	.11	.22	.38	.58	.21	.53	.85	.98
$GLS - WS$	.12	.23	.40	.59	.20	.51	.80	.94
$GLS_u - WS$	.12	.23	.40	.60	.21	.54	.86	.98
$REC - WS$	.07	.17	.33	.54	.15	.47	.82	.97
$GLS - MAX$	.11	.22	.39	.59	.21	.52	.81	.94
$GLS_u - MAX$	.11	.22	.38	.58	.20	.53	.85	.98
$REC - MAX$	.05	.15	.32	.54	.13	.46	.82	.97

Table 8. Power of tests at nominal 0.05-level, linear trend, $p = 2$, DGP: $(1 - \rho L)^2 y_t = \epsilon_t$.

	$T = 50$				$T = 100$			
ρ	.90	.80	.60	.40	.90	.80	.70	.60
DF	.07	.10	.24	.54	.10	.22	.46	.74
GLS	.07	.11	.30	.63	.11	.29	.56	.80
GLS_u	.07	.10	.29	.63	.11	.28	.56	.83
REC	.03	.06	.24	.60	.05	.20	.49	.79
WS	.07	.11	.32	.67	.11	.29	.58	.85
MAX	.07	.11	.31	.65	.11	.28	.57	.84
$GLS - WS$	.08	.11	.31	.63	.11	.29	.56	.80
$GLS_u - WS$	.08	.12	.32	.66	.11	.29	.58	.84
$REC - WS$	.03	.06	.24	.61	.05	.20	.49	.79
$GLS - MAX$	.07	.11	.30	.61	.11	.28	.55	.79
$GLS_u - MAX$	.07	.11	.30	.65	.11	.29	.57	.84
$REC - MAX$	.02	.05	.22	.60	.04	.18	.47	.79

Table 9. Size of tests at nominal 0.05-level, $p = 6$, DGP: $(1 - L)y_t = \epsilon_t$.

	constant		linear trend	
	$T = 50$	$T = 100$	$T = 50$	$T = 100$
DF	.045	.047	.041	.045
GLS	.037	.043	.030	.043
GLS_u	.029	.038	.024	.037
REC	.027	.035	.018	.026
WS	.051	.049	.048	.050
MAX	.042	.046	.040	.043
$GLS - WS$	.043	.042	.031	.041
$GLS_u - WS$	.049	.047	.035	.043
$REC - WS$	.029	.035	.020	.027
$GLS - MAX$	.029	.036	.022	.037
$GLS_u - MAX$	.039	.043	.026	.038
$REC - MAX$	.026	.035	.009	.022

Table 10. Size of tests at nominal 0.05-level using asymptotic critical values, $p = 1$, DGP: $(1 - L)y_t = \epsilon_t$.

	constant		linear trend	
	$T = 50$	$T = 100$	$T = 50$	$T = 100$
DF	.053	.051	.062	.053
GLS	.133	.090	.087	.067
GLS_u	.058	.055	.069	.056
REC	.045	.047	.042	.041
WS	.055	.053	.072	.060
MAX	.050	.050	.059	.054
$GLS - WS$	.099	.082	.086	.067
$GLS_u - WS$	.054	.053	.069	.058
$REC - WS$	.047	.048	.045	.042
$GLS - MAX$	.179	.141	.106	.075
$GLS_u - MAX$	.053	.052	.067	.058
$REC - MAX$	.045	.047	.042	.043

Table 11. Size of tests at nominal 0.05/0.10 level, constant only, $p = 1$
DGP: $(1 - L)y_t = \epsilon_t$
for non-normal innovations ϵ_t; $T = 50$.

Innovation distribution	$\chi^2(1) - 1$		t_5	
Nominal size	.050	.100	.050	.100
DF	.042	.087	.050	.089
GLS	.045	.094	.044	.096
GLS_u	.044	.087	.051	.085
REC	.043	.092	.044	.092
WS	.043	.094	.043	.098
MAX	.043	.093	.043	.097
$GLS - WS$	.045	.094	.044	.098
$GLS_u - WS$	.045	.094	.043	.098
$REC - WS$	.043	.092	.044	.094
$GLS - MAX$	.044	.093	.044	.096
$GLS_u - MAX$	.044	.093	.043	.098
$REC - MAX$	.045	.095	.043	.100