New Unit Root Tests with Nonnormal Errors

Kyung So Im
Federal Deposit Insurance Corporation

Junsoo Lee
University of Alabama

Margie Tieslau
University of North Texas

August 31, 2009
(not complete; do not quote)

Abstract

This paper proposes new unit root tests which are more powerful when the error term follows a non-normal distribution. The improved power is gained by utilizing more moment conditions through a computationally simple procedure. Specifically, we follow the residual augmented least squares (RALS) estimator proposed by Im and Schmidt (2008) in order to use the information implied by nonnormal errors. Our Monte Carlo simulation results show that the size of the RALS-based unit root tests is quite close to the asymptotic size and the power is improved significantly over the usual Dickey-Fuller tests when the error is not normal. As such, our finding shows significant efficiency gain when the information on nonnormality is utilized, although this information has been ignored in usual unit root tests.

JEL Classification: C22, C12, C13.

Key Words: Unit root test, Generalized methods of moments, Residual augmented least squares.

1 Introduction

The usual unit root tests are based on least squares estimation. Using least squares estimation yields desirable properties. The so-called Gauss Markov theorem indicates that the least squares estimator is the most efficient under the classical assumptions which include the normal distribution of the error term. Conventional wisdom says that when the distribution of the error term is not normal, this will not pose a serious problem. The distribution of usual unit root tests will not be affected significantly with nonnormal errors. As such, nonnormality of the error term has been often ignored in the usual unit root tests. However, this outcome does not imply that we cannot utilize the information embodied in nonnormal errors. We wonder if more powerful tests can be developed when the information on nonnormality is utilized. This issue seems important since low power has often been cited as a drawback in the usual unit root tests. A nonnormal distribution in the error term can entail useful information. For example, it is well known that many financial data have fat tailed distributions, a skewed distribution, a leptokurtic distribution, or even a mixture of two different distributions. Nonetheless, the usual unit root tests using linear least squares estimation ignore such information. Nonnormal distributions of the error term can occur for a variety of reasons and may not be easily distinguishable from forms of nonlinearity. For instance, a mixture of two distributions is also a feature of regime switching models. Skewed distributions can occur when an asymmetric relationship exists in the data and threshold models are often applied in such cases. A fat tailed distribution is frequently observed in high frequency time series data and it is modeled in a nonlinear model framework including ARCH models. However, we do not know exactly which nonlinear model will fit the best.

A few authors have previously investigated the possibility of utilizing the information contained in nonnormal errors. For example, Cox and Llatas (1991) studied the asymptotic distribution of maximum likelihood estimators (MLE) in the Dickey-Fuller regression assuming that the true error density is known. Lucas (1995) derived the asymptotic distribution of the unit root test statistic based on the M-estimator. However, it is assumed that the distribution of the error term is known, but the distribution is usually unknown.

In this paper we follow the framework of generalized methods of moments (GMM) and seek to find more powerful tests by utilizing moment conditions implied by non-normal errors. Unlike Cox and Llatas (1991) and others, it is not necessary to specify the particular density function of the error term. We first derive the asymptotic distribution of the GMM-based unit root test. Then, following Im and Schmidt (2008), we propose a simplified two-step procedure, which is referred to as "residual augmented least squares" (RALS), and show that the distribution of the RALS-based test is asymptotically identical to the asymptotic distribution of GMM in testing the unit root hypothesis. Our suggested procedure is simple and easy. We do not need to specify an unknown nonlinear model or unknown density function. Instead, we suggest utilizing the information on nonnormal errors from the residuals of usual augmented Dickey-Fuller test regression, and use it as additional moment conditions. Our work is extended to the unit root testing procedure using the results in Im and Schmidt (2008) that showed that the RALS estimator is asymptotically identical to the GMM estimator. The power gain is significant when the errors are not normal. Specifically, we find that the power of the RALS-based test is far better than that of the ADF test. Our simulation results also show that the small sample size of the RALS-based unit root test is quite close to the asymptotic size. Thus, we can achieve improved powers without inducing size distortions.

The rest of the paper is organized as follows. In Section 2, we derive the asymptotic distribution of the GMM based unit root tests. In Section 3, we propose RALS-based unit root tests and provide the asymptotic distribution when the errors have nonnormal distributions. In Section 4, we provide an empirical example. Section 5 provides concluding remarks.

2 Asymptotic Distribution of GMM Unit Root Statistics

In this section we derive the asymptotic distributions of GMM estimators resulting from least squares moments when some additional moment restrictions are used. We also derive the asymptotic distributions of their associated t-statistics.

Consider a time series that follows:

$$y_t = \phi y_{t-1} + \varepsilon_t,\ t = 1, 2, ... \tag{1}$$

where $\{\varepsilon_t\}_{t=1}^{\infty}$ is a sequence of innovations. We are interested in the test of the unit root hypothesis $H_0 : \phi = 1$ against the alternative hypothesis $H_A : \phi < 1$. We assume:

Assumption 1. $\varepsilon_t = \sum_{j=1}^{p} a_j \varepsilon_{t-j} + e_t,\ t = 1, 2, ...,$ where $\{e_t\}_{t=1}^{\infty}$ is an iid sequence with zero mean and finite second moment σ_e^2, and all roots of $a(z) = 1 - \sum_{j=1}^{p} a_j z^j$ lie outside of the unit circle.

If Assumption 1 is met and $y_0 = 0$, an appropriate model is ADF:

$$\Delta y_t = \beta y_{t-1} + \sum_{j=1}^{p} \delta_j \Delta y_{t-j} + e_t,\ t = 1, 2, ..., \tag{2}$$

where $\Delta y_t = y_t - y_{t-1}$ and $\beta = \phi - 1$. Let $\hat{\beta}_{LS}$ be the least squares estimator of β in regression (2), and let t_{LS} be its associate t-statistic. Then, it is well known, under the null hypothesis, that

$$T\hat{\beta}_{LS} \Rightarrow a(1)\left(\int_0^1 [W(r)]^2\, dr\right)^{-1} \int_0^1 W(r)dW(r), \tag{3}$$

and

$$t_{LS} \Rightarrow \left(\int_0^1 [W(r)]^2 dr\right)^{-1/2} \int_0^1 W(r)dW(r) \equiv DF, \tag{4}$$

where $a(1) = 1 - \sum_{j=1}^{p} a_j$, and $W(r)$ denotes the standard Brownian motion on $r \in [0, 1]$.

Let $\xi_t = (\Delta y_{t-1}, \Delta y_{t-2}, ..., \Delta y_{t-p})'$, and $z_t = (y_{t-1}, \xi_t')'$. Suppose we have $J \times (p + 1)$ additional moment conditions:

$$E\left[g(e_t) \otimes z_t\right] = 0,\ t = 1, 2, ..., \tag{5}$$

where $g(e_t)$ is a $J \times 1$ vector that satisfies the following:

Assumption 2. $g(\cdot)$ is differentiable and satisfies the first-order Lipshitz condition: $\left|g_j'(x) - g_j'(y)\right| < M\left|x - y\right|$ for some constant M for all j, where $g_j(\cdot)$ is the j-th element of $g(\cdot)$. $E\left[g(e_t)\right] = 0$, the second moment of $g(e_t)$ exists, and $E\left[g'(e_t)\right] < \infty$.

Define $C = E[g(e_t)g(e_t)']$ and $D = E[g'(e_t)]$, and $\psi(e_t) = D'C^{-1}g(e_t)$, for $t = 1, 2, ...$ Also define the correlation between e_t and $\psi(e_t)$ by

$$\rho = \frac{\sigma_{\psi e}}{\sigma_\psi \sigma_e} \tag{6}$$

where $\sigma_\psi^2 = Var[\psi(e_t)] = Var[D'C^{-1}g(e_t)] = D'C^{-1}D$, $\sigma_{\psi e} = E[\psi(e_t)e_t] = DC^{-1}E[g(e_t)e_t]$.

Theorem 1. Suppose a time series follows (1), and Assumptions 1 and 2 are satisfied. Under the null hypothesis, we have, for, $\tilde{\beta}_G$, the GMM estimator using the moments conditions (5) in the ADF regression (2):

$$T\tilde{\beta}_G \Rightarrow \frac{a(1)}{\sigma_e \sigma_\psi}\left(\int_0^1 [W_1(r)]^2 dr\right)^{-1} \int_0^1 W_1(r)dW_2. \tag{7}$$

Also, for its t-statistic obtained by $t_G = \tilde{\beta}_G / se(\tilde{\beta}_G)$, where:

$$se(\tilde{\beta}_G) = \tilde{\sigma}_\psi^{-1}\sqrt{\left(\sum_{t=1}^T y_{t-1}^2 - \sum_{t=1}^T y_{t-1}\xi_t \left(\sum_{t=1}^T \xi_t \xi_t'\right)^{-1} \sum_{t=1}^T \xi_t' y_{t-1}\right)^{-1}},$$

$\tilde{\sigma}_\psi^2 = \tilde{D}'\hat{C}^{-1}\tilde{D}$, $\tilde{D} = T^{-1}\sum_{t=1}^T g'(\tilde{e}_t)$, $\tilde{C} = T^{-1}\sum_{t=1}^T g(\tilde{e}_t)g(\tilde{e}_t)'$, and $\tilde{e}_t$ is the residual from GMM estimation in the regression (2), we then have:

$$t_G \Rightarrow \rho DF + \sqrt{1-\rho^2}N(0,1), \tag{8}$$

where ρ is defined in (6), DF denotes the Dickey-Fuller distribution as was defined in (4), and $N(0,1)$ signifies the standard normal distribution.

proof. See Appendix.

In the case where an intercept is allowed in the regression (2), we have the model:

$$\Delta y_t = \alpha_1 + \beta y_{t-1} + \sum_{j=1}^p \delta_j \Delta y_{t-j} + e_t, \ t = 1, 2, ..., \tag{9}$$

and we have the additional moment conditions $E\left[g(e_t)\otimes(1,z_t)'\right]=0$. In view of the expression for the estimator of β in (A.9) of the Appendix, this produces the GMM estimator given by:

$$T\tilde{\beta}_{G,\mu}=\left(\sigma_\psi^2 T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^2\right)^{-1}T^{-1}\sum_{t=1}^{T}\tilde{y}_{t-1}\psi(e_t)+o_p\left(1\right),$$

where $\tilde{y}_{t-1}=y_{t-1}-T^{-1}\sum_{t=1}^{T}y_{t-1}$, $t=1,2,....$ Consequently,

$$T\tilde{\beta}_{G,\mu}\Rightarrow\frac{a(1)}{\sigma_\psi\sigma_e}\int_0^1\tilde{W}_1(r)dW_2(r)/\int_0^1\left[\tilde{W}_1(r)\right]^2dr, \tag{10}$$

where $\tilde{W}_1(r)$ is the demeaned Brownian motion: $\tilde{W}_1(r)=W_1(r)-\int_0^1 W_1(r)dr$. Also, by construction:

$$t_{G,\mu}\Rightarrow\rho DF_\mu+\sqrt{1-\rho^2}N(0,1),$$

where DF_μ denotes the limiting distribution of the t-statistic from least squares in regression (9).

Similarly, when the model includes a linear time trend as well as an intercept, we have the model:

$$\Delta y_t=\alpha_1+\alpha_2 t+\beta y_{t-1}+\sum_{j=1}^{p}\delta_j\Delta y_{t-j}+e_t,\ t=1,2,..., \tag{11}$$

and this will result in the GMM estimator given by:

$$T\tilde{\beta}_{G,\tau}\Rightarrow\frac{a(1)}{\sigma_\psi\sigma_e}\int_0^1\breve{W}_1(r)d\breve{W}_2(r)/\int_0^1\left[\breve{W}_1(r)\right]^2dr, \tag{12}$$

where $\breve{W}(r)$ is the detrended Brownian motion. Also,

$$t_{G,\tau}\Rightarrow\rho DF_\tau+\sqrt{1-\rho^2}N(0,1), \tag{13}$$

where DF_τ denotes the limiting distribution of the t-statistic for the OLS estimator of β in the regression (11).

Remark 1. The asymptotic distribution of t_G depends on the nuisance parameter ρ. Hansen (1995) reports the critical values of the asymptotic distribution of the t-statistics for $\rho^2=0.1$ to 1.0 at intervals of 0.1.

3 RALS Estimation

Consider the model (9), the ADF model with an intercept. Suppose $g(e_t) = \left(e_t, [h(e_t) - K]'\right)'$. Let $x_t = (1, z_t')'$. Then, we have the moment condition $E[g(e_t) \otimes x_t] = 0$, which we split into two parts: the least squares moment conditions given by

$$E(e_t \otimes x_t) = 0 \tag{14}$$

and an additional $2(J-1)$ moment conditions given by

$$E[(h(e_t) - K) \otimes x_t] = 0. \tag{15}$$

Therefore, we have

$$C = \begin{bmatrix} \sigma_e^2 & C_{21}' \\ C_{21} & C_{22} \end{bmatrix}, \text{ and } D = \begin{bmatrix} 1 \\ D_2 \end{bmatrix}, \tag{16}$$

where $C_{21} = E[e_t h(e_t)]$, $C_{22} = E[h(e_t)h(e_t)']$, and $D_2 = E[h'(e_t)]$.

Let

$$\hat{w}_t = h(\hat{e}_t) - \hat{K} - \hat{e}_t \hat{D}_2, \; t = 1, 2, ..., \tag{17}$$

where $\hat{e}_t$ is the OLS residual from the regression (9), $\hat{K} = \frac{1}{T}\sum_{t=1}^{T} h(\hat{e}_t)$, $\hat{D}_2 = \frac{1}{T}\sum_{t=1}^{T} h'(\hat{e}_t)$. The RALS-based testing equation is given by:

$$\Delta y_t = \alpha_1 + \beta y_{t-1} + \sum_{j=1}^{p} \delta_j \Delta y_{t-j} + \hat{w}_t' \gamma + \eta_t, \; t = 1, 2, ... \tag{18}$$

In the following, we show that the RALS estimator is asymptotically identical to the GMM estimator using moment conditions (14) and (15).

Theorem 2. Consider a time series as described in (1) with $\phi = 1$. Under Assumptions 1 and 2, the RALS estimator of β, which is obtained by applying least squares to (18), is asymptotically identical to the GMM estimator using the moment conditions (14) and (15). In addition, the limiting distributions of the t-statistics are the same.

proof. See the Appendix.

When there is a linear time trend included in the regression, we have:

$$\Delta y_t = \alpha_1 + \alpha_2 t + \beta y_{t-1} + \sum_{j=1}^{p} \delta_j \Delta y_{t-j} + \hat{w}_t' \gamma + \eta_t, \;\; t = 1, 2, ... \tag{19}$$

By construction, we will obtain the same results as those in (12) and (13) for the estimator of β and for its t-statistic.

Next, we provide some guidance on how to apply the RALS procedure in practice.

- ρ^2 is estimated by

  $$\hat{\rho}^2 = \hat{\sigma}_A^2 / \hat{\sigma}^2,$$

  where $\hat{\sigma}^2$ is the usual estimate of the error variance in the standard ADF regression, and $\hat{\sigma}_A^2$ is the estimate of the error variance in the RALS regression. See the proof of Theorem 2 [equations (A.16) and (A.19)]. Based on this value of $\hat{\rho}^2$, the critical values are found in Hansen (1995).

- When the sample size is small (e.g. $T \leq 50$), impose the restriction of $\beta = 0$ in the first step regression that yields the residuals for the augmented variables in $\hat{w}_t$. According to our simulations, this procedure significantly improves the size property of the test with only minimal effects on power. When the sample is relatively big (e.g., $T = 100$), this effect, however, disappears quickly.

4 Simulation Results

In this section, we investigate the small sample property of the unit root tests based on RALS. As was noted above, we use the critical values reported by Hansen (1995) to evaluate the RALS-based test, and we impose the restriction $\beta = 0$ when we construct the augmented variable. For example, suppose we have a regression model: $\Delta y_t = \alpha + \beta y_{t-1} + \sum_{j=1}^{p} \delta_j \Delta y_{t-j} + e_t$, $t = 1, 2, ...$ In the first step, we estimate α and δ' imposing $\beta = 0$ so that we begin with the regression $\widehat{\Delta y}_t = \hat{\alpha} + \sum_{j=1}^{p} \hat{\delta}_j \Delta y_{t-j}$ and then construct the augmented variable $\hat{w}_t$ as a function of the residuals $\hat{e}_t = \Delta y_t - \hat{\alpha} - \sum_{j=1}^{p} \hat{\delta}_j \Delta y_{t-j}$. Then, in the second step, the t-statistic is computed following the usual procedure from the RALS regression: $\Delta y_t = \tilde{\alpha} + \tilde{\beta} y_{t-1} + \sum_{j=1}^{p} \tilde{\delta}_j \Delta y_{t-j} + \tilde{\gamma} \hat{w}_t + \hat{v}_t$.

We include two RALS estimators, RALS(2&3) and RALS(t5), in our Monte Carlo study. First, RALS(2&3) imposes the additional moment conditions that the second and third moments of the errors are not correlated with the lagged dependent variables. Therefore, $h(\hat{e}_t) = [\hat{e}_t^2, \hat{e}_t^3]'$. Letting $m_j = T^{-1} \sum_{t=1}^{T} \hat{e}_t^j$, for $j = 2, 3$, we have for RALS(2&3):

$$\hat{w}_t = \left[\hat{e}_t^2 - m_2,\ \hat{e}_t^3 - m_3 - 3m_2 \hat{\varepsilon}_t\right]',\ t = 1, 2, ... \qquad (20)$$

The moment condition $E\left[\left(e_t^2 - \sigma_e^2\right) y_{t-1}\right] = 0$ is the condition of no heteroskedasticity. This condition improves the efficiency of the estimator of β when the errors are not symmetric. The restriction on the third moments conditional on y_{t-1} improves efficiency unless $\mu_4 = 3\sigma^4$. See Im and Schmidt (2008) for details.

Second, RALS(t5) imposes the restrictions that arise from the score of the maximum likelihood procedure when the error density is assumed to be a t-distribution with 5 degrees of freedom. The assumption of a student-t error density is a widely accepted strategy in the M-estimate and this is known to lead to more efficient estimation than OLS when the true density has fat-tails. Because RALS(t5) also uses the least squares restrictions, RALS(t5) would be as efficient as least squares when the error is normally distributed, and is more efficient when the density of the errors have fat-tails. In this case, we have $h(e_t) = (c+1)\, e_t/(c+e_t^2)$, so $D_2 = (c+1)\left(c - e_t^2\right)/(c+e_t^2)^2$, where $c = 5$. Therefore, in this scenario we have:

$$\hat{w}_t = \frac{6\hat{e}_t}{5+\hat{e}_t^2} - \frac{1}{T}\sum_{t=1}^{T} \frac{6\hat{e}_t}{5+\hat{e}_t^2} - \hat{e}_t \frac{1}{T}\sum_{t=1}^{T} \frac{6\left(5-\hat{e}_t^2\right)}{(5+\hat{e}_t^2)^2} \qquad (21)$$

There is no compelling reason behind choosing $c = 5$. However, it seems that the tests are quite robust to the selection of different values of c. For example, our simulations that use $c = 3$, which are not reported here for space considerations, indicate that the empirical size and power of the tests are almost identical to the case when $c = 5$.

To examine the size property , we report the rejection ratio for $\alpha = 0.05$ when $\phi = 1$; to examine the power, we use $\phi = 0.9$. We simulated the sample cases for $T = 50$ and 100. All the results are based on 5,000 replications.

Table 1 reports the results for the basic case when the errors, ε_t, in (1) are serially independent and p, the number of ADF augmentation terms in the regression, is set to zero. We compare RALS(2&3) and RALS(t5) to three competing test statistics: DF, the standard Dickey-Fuller test based on OLS; AD, the test studied by Beelders (1996) and Shin and So (1999) based on adaptive estimation; and M5, the test based on the M-estimate assuming the true density is student-t with 5 degrees of freedom, which was studied by Lucas (1995). The figures for AD and M5 have been reproduced from Shin and So (1999). We replicated the four distributions simulated by Shin and So (1999): (i) standard normal, (ii) t-distribution with $df = 3$, (iii) mixture normal: 0.5N(-3,1)+0.5N(3,1), and (iv) chi-square with $df = 1$.

As is seen in Table 1, the sizes of the tests based on RALS(2&3) and RALS(t5) are quite close to the nominal 5% throughout, and the power gain over the standard DF test is substantial when the errors are not normal. The overall power of RALS(2&3) and RALS(t5) compares favorably to the power of AD or M5. The performance of RALS(t5) and M5 are similar when the true density is the student-t with 3 degrees of freedom, but RALS(t5) is better when the density is mixture normal. When the true density is a chi-square distribution with one degree of freedom, RALS(2&3), which explicitly uses the moment condition that is useful when the error is not symmetric, dominates the other tests. The AD-based test does not seem to capture the possible efficiency gain from the non-symmetric feature of the error density. In our simulated distributions, RALS(t5) is marginally better than RALS(2&3) when the density is symmetric. However, as we can see for the case when the density is chi-square with one degree of freedom, RALS(2&3) is generally better than RALS (t5) when the error density is skewed, and the difference often is quite substantial.

In Tables 2-5, we compare the performance of the tests when the errors are serially correlated. In doing so, we compare only three tests, ADF, RALS(2&3) and RALS(t5), in two data generation processes:

$$AR : \varepsilon_t = 0.5\varepsilon_{t-1} + e_t,\ t = 1, 2, ...,$$

and

$$MA : \varepsilon_t = e_t - 0.5e_{t-1},\ t = 1, 2, ...$$

We report the size and power for fixed ADF augmentation at $p = 2$ and $p = 4$ when $T = 50$, and $p = 3$ and $p = 6$ when $T = 100$, as well as when p is selected by information criterion. We simulated the Akaike and Schwarz criteria, but report only the results from the Schwarz criterion since the results from the Akaike criterion were similar. The minimum and maximum values of p are set 2 and 4 when $T = 50$, and 3 and 6 when $T = 100$. We consider the case when the errors are generated from the standard normal, Cauchy, student-t distribution with 2 degrees of freedom, double exponential, chi-square distribution with 4 degrees of freedom, and beta(2,2) distribution. The Cauchy and the t-distribution with 2 degrees of freedom do not satisfy Assumptions 1 and 2, so we do not know the asymptotic distributions of the statistics in this case. However, it is interesting to see the performance of the tests in this situation.

Table 2 reports the size and power of the tests for AR(1) errors when a time trend is not included in ADF regressions, and Table 3 contains the case when a linear time trend is allowed. The sizes of all of the three tests reported both in Tables 2 and 3 are close to the 5% nominal size, in general, even when the errors are generated from a Cauchy or t-distribution with two degrees of freedom. The only exception is RALS(t5). The empirical size of the RALS(t5)-based test is 10-12% when the errors are generated from a Cauchy and a time trend is allowed in the ADF regression. This result is somewhat puzzling, especially since the empirical size of the RALS(t5)-based test is close to the nominal size when there is no time trend included in the regression.

The power difference between the two tests based on OLS and RALS is the greatest when the errors are generated from a Cauchy distribution. Also, as we observed in Table 1, RALS(t5) is more powerful than RALS(2&3) for all the symmetric distributions. However, RALS(2&3) is, in general, powerful when the errors are asymmetric. In particular, it is seen that the power of the RALS(t5)-based test is lower than that of the OLS-based tests when the error is chi-square distributed with 4 degrees of freedom; the power of RALS(2&3) is 52% while the power of RALS(t5) is 15% (see Table 3) when a time trend is included, $T = 100$ and $p = 3$.

Tables 4 and 5 contain the results when the errors follow an MA(1) model. When p

is determined by the Schwarz criteria, all the tests tend to over-reject the null hypothesis. However, when p is fixed at 4 for $T = 50$, and at $p = 6$ for $T = 100$, the size of the tests are quite close to the 5% nominal size, except when the density is Cauchy and the regression includes time trend. But, the overall size of RALS(2&3) seems as robust as the size of the standard ADF test. With regard to the power of the tests, we observe a similar pattern as in the case of AR(1) errors. RALS-based tests are substantially more powerful than OLS-based ADF tests, and RALS(2&3) compares favorably to RALS(t5).

5 An Application of the RALS Unit Root Test

In this section, we present an example of the RALS unit root test applied to 14 key US macroeconomic variable analyzed by Nelson and Plosser (1982). These variables include real and nominal GNP, real per capita GNP, the index of industrial production, total employment, the unemployment rate, the GNP deflator, the consumer price index, nominal and real wages, the M1 money stock, M1 velocity of money, bond yields and stock prices. Nelson and Plosser's results indicated that all variables except for unemployment were difference stationary. We re-examine this issue by applying our test to an updated version of the Nelson & Plosser data set, comparing the results of our test to those of the usual ADF test applied to each series. (more to come)

6 Concluding Remarks

This paper proposes new unit root tests which are more powerful when the error term follows a non-normal distribution. The improved power is gained by utilizing more moment conditions through a computationally simple procedure. Specifically, we follow the residual augmented least squares (RALS) estimator proposed by Im and Schmidt (2008) in order to use the information implied by nonnormal errors. Such a simple procedure becomes available as the asymptotic distribution of GMM estimators is shown to be the same as that from a simple RALS-based testing procedure. Our Monte Carlo simulation results show that the size of the RALS-based unit root tests is quite close to the asymptotic size and the power is improved significantly over the usual Dickey-Fuller tests when the error is not normal. As such, our finding shows significant efficiency gain when the information on nonnormality is utilized, although this information has been ignored in usual unit root tests.

A Appendix

Lemma A1. $z_t = (y_{t-1}, \xi_t')'$, as was defined previously in equation (5), and C and D are as defined in (16). Let the $(p+1) \times (p+1)$ matrix $\Upsilon_T = diag\left(T, \sqrt{T}, ..., \sqrt{T}\right)$. Then, we have, under Assumptions 1 and 2, and under the null hypothesis,

$$\sum_{t=1}^{T} \left[g'(e_t) \otimes \Upsilon_T^{-1} z_t z_t' \Upsilon_T^{-1}\right] \Rightarrow D \otimes \int zz', \tag{A.1}$$

$$\sum_{t=1}^{T} g(e_t) g(e_t)' \otimes \Upsilon_T^{-1} z_t z_t' \Upsilon_T^{-1} \Rightarrow C \otimes \int zz', \tag{A.2}$$

where $\int zz' = diag\left(a(1)^{-2}\sigma_e^2 \int_0^1 \left[W_1(r)\right]^2 dr,\ E\left(\xi_t \xi_t'\right)\right)$. Also:

$$\sum_{t=1}^{T} \psi(e_t) \Upsilon_T^{-1} z_t = \begin{bmatrix} T^{-1} \sum_{t=1}^{T} \psi(\varepsilon_t) y_{t-1} \\ T^{-1/2} \sum_{t=1}^{T} \psi(\varepsilon_t) \xi_t \end{bmatrix} \Rightarrow \begin{bmatrix} \frac{\sigma_\psi \sigma_\varepsilon}{a(1)} \int_0^1 W_1(r) dW_2(r) \\ \Gamma \end{bmatrix}, \tag{A.3}$$

where $[W_1(r), W_2(r)]'$ is a bivariate Brownian motion with correlation ρ, and Γ is $p \times p$ multivariate normal variable with covariance matrix $\sigma_\psi^2 E\left(\xi_t \xi_t'\right)$.

proof. Lucas (1995, Lemma 1 in Appendix). See also Hansen (1995, Lemma).

Lemma A2. ρ is defined as in equation (6). Then,

$$\rho = \frac{1}{\sigma_e \sigma_\psi}. \tag{A.4}$$

Also,

$$\frac{1}{\sigma_\psi^2} = \sigma_e^2 - \left(C_{21} - \sigma_e^2 D_2\right)' \left(C_{22} + \sigma_e^2 D_2 D_2' - C_{21} D_2' - D_2 C_{21}'\right)^{-1} \left(C_{21} - \sigma_e^2 D_2\right). \tag{A.5}$$

proof. The first result follows from routine matrix algebra using the partitioned inverse lemma. For the second result, we have, after straightforward algebra,

$$\left(D' C^{-1} D\right)^{-1} = \sigma_e^2 \left(1 + \left(C_{21} - \sigma_e^2 D_2\right)' \left(\sigma_e^2 C_{22} - C_{21} C_{21}'\right)^{-1} \left(C_{21} - \sigma_e^2 D_2\right)\right)^{-1},$$

which, however, is the same as $1/\sigma_\psi^2$ from Amemiya (1985, p461, Lemma 20).

PROOF OF THEOREM 1: We note the entire proof follows immediately from Lucas (1995, Theorem 1) since the GMM estimator is obtained by solving the score $\sum_{t=1}^{T}\left[DC^{-1}g(e_t)z_t\right] = \sum_{t=1}^{T}\left[\psi(e_t)z_t\right] = 0$, and this score could be thought as that of the M-estimate. However, we provide more details. Let $\theta = (\beta, \delta_1, \delta_2, ..., \delta_p)'$. The GMM estimator is obtained by solving:

$$\min_{\theta} \sum_{t=1}^{T}\left[g(e_t)\otimes z_t\right]' \hat{\Lambda}^{-1} \sum_{t=1}^{T}\left[g(e_t)\otimes z_t\right], \tag{A.6}$$

where $\hat{\Lambda} = \left(\sum_{t=1}^{T} g(\hat{e}_t)g(\hat{e}_t)' \otimes z_t z_t'\right)$, and $\hat{e}_t$ is the residual from an initial consistent estimator of θ. Taking the derivative with respect to θ, we obtain the score:

$$\sum_{t=1}^{T}\left[g'(\tilde{e}_t)\otimes z_t z_t'\right]' \hat{\Lambda}^{-1} \sum_{t=1}^{T}\left[g(\tilde{e}_t)\otimes z_t\right] = 0, \tag{A.7}$$

where $\tilde{e}_t = \Delta y_t - z_t\tilde{\theta}$, and $\tilde{\theta}$ is the GMM estimator. Note that $\Upsilon_T = diag\left(T, \sqrt{T}, ..., \sqrt{T}\right)$. The Taylor series expansion of the term $\sum_{t=1}^{T}\left[g(\tilde{e}_t)\otimes z_t\right]$ with respect to the true disturbance e_t and premultiplication of $I_J \otimes \Upsilon_T^{-1}$ yields:

$$\begin{aligned}&\sum_{t=1}^{T}\left[g(\tilde{e}_t)\otimes \Upsilon_T^{-1}z_t\right] \\ =\ &\sum_{t=1}^{T}\left[g(e_t)\otimes \Upsilon_T^{-1}z_t - g'(e_t)\otimes \Upsilon_T^{-1}z_t z_t'\Upsilon_T^{-1}\Upsilon_T\left(\tilde{\theta}-\theta\right)\right] + o_p(1).\end{aligned} \tag{A.8}$$

Solving (A.7) with respect to $\Upsilon_T\left(\tilde{\theta}-\theta\right)$, after substituting (A.8) into (A.7), we obtain:

$$\begin{aligned}&\Upsilon_T(\tilde{\theta}-\theta) = \\ &\left\{\sum_{t=1}^{T}\left[g'(\tilde{e}_t)\otimes \Upsilon_T^{-1}z_t z_t'\Upsilon_T^{-1}\right]'\left[\sum_{t=1}^{T} g(\hat{e}_t)g(\hat{e}_t)'\otimes \Upsilon_T^{-1}z_t z_t'\Upsilon_T^{-1}\right]^{-1}\sum_{t=1}^{T}\left[g'(e_t)\otimes \Upsilon_T^{-1}z_t z_t'\Upsilon_T^{-1}\right]\right\}^{-1} \\ &\times\left\{\sum_{t=1}^{T}\left[g'(\tilde{e}_t)\otimes \Upsilon_T^{-1}z_t z_t'\Upsilon_T^{-1}\right]'\left[\sum_{t=1}^{T} g(\hat{e}_t)g(\hat{e}_t)'\otimes \Upsilon_T^{-1}z_t z_t'\Upsilon_T^{-1}\right]^{-1}\sum_{t=1}^{T}\left[g(e_t)\otimes \Upsilon_T^{-1}z_t\right]\right\} + o_p(1).\end{aligned} \tag{A.9}$$

Noting that

$$\sum_{t=1}^{T}\left\{\left[g'(\tilde{e}_t) - g'(e_t)\right]\otimes \Upsilon_T^{-1}z_t z_t'\Upsilon_T^{-1}\right\} = o_p(1),$$

and

$$\sum_{t=1}^{T}\left\{\left[g(\hat{e}_t)g(\hat{e}_t)' - g(e_t)g(e_t)'\right]\otimes \Upsilon_T^{-1}z_t z_t'\Upsilon_T^{-1}\right\} = o_p(1),$$

we have, from Lemma A1:

$$T\tilde{\beta}_G \Rightarrow \frac{a(1)}{\sigma_\psi \sigma_e}\left(\int_0^1 [W_1(r)]^2 dr\right)^{-1}\int_0^1 W_1(r)dW_2(r), \tag{A.10}$$

where $[W_1(r), W_2(r)]$ is a bivariate Brownian motion with correlation ρ. We have for the t-statistic:

$$t_G \Rightarrow \left(\int_0^1 [W_1(r)]^2 dr\right)^{-1/2}\int_0^1 W_1(r)dW_2(r), \tag{A.11}$$

which is a mixture of the Dickey-Fuller and standard normal distribution described in (8). To see this, note that:

$$T^{-1/2}\sum_{t=1}^{[rT]}\begin{bmatrix} e_t \\ \psi(e_t) \end{bmatrix} \Rightarrow \begin{bmatrix} \sigma_e W_1(r) \\ \sigma_\psi W_2(r) \end{bmatrix}, \tag{A.12}$$

where $[rT]$ denotes the integer part of rT. Therefore:

$$W_2(r) = \rho W_1(r) + \sqrt{1-\rho^2}W_3(r), \tag{A.13}$$

where $W_3(r)$ is independent of $W_1(r)$. The result follows if we note that

$$\left(\int_0^1 [W_1(r)]^2 d(r)\right)^{-1/2}\int_0^1 W_1(r)dW_3(r)$$

is standard normal.

PROOF OF THEOREM 2: Define a variable as a function of true disturbances:

$$w_t = h(e_t) - K - e_t D_2,\ t = 1, 2, ...$$

The variables in w_t are not observable, but we momentarily assume that they are observed. Then we show that the augmentation of w_t or $\hat{w}_t$ asymptotically yields the same estimator of $T\beta$. Consider a regression:

$$\Delta y_t = \alpha_1 + \beta y_{t-1} + \sum_{j=1}^{p}\delta_j \Delta y_{t-j} + w_t'\gamma + v_t,\ t = 1, 2, ... \tag{A.14}$$

Therefore,

$$e_t = w_t'\gamma + v_t,\ t = 1, 2, ... \tag{A.15}$$

Let $\hat{\beta}_A^*$ be the least squares estimator of β from regression (A.14), $\sigma_v^2 = Var(v_t)$, and

$$\lambda = \frac{\sigma_{ev}}{\sigma_e \sigma_v} = \frac{\sigma_v}{\sigma_e}, \tag{A.16}$$

where $\sigma_{ev} = E(\varepsilon_t v_t)$. The second equality of (A.16) follows since w_t and v_t are not correlated, so that $\sigma_{ev} = \sigma_v^2$. From Hansen (1995, Theorem 2 and 3), we have:

$$T\hat{\beta}_A^* \Rightarrow \frac{\sigma_v}{\sigma_e}\left(\int_0^1 [W_4(r)]^2\right)^{-1}\int_0^1 W_4(r)dW_5(r), \tag{A.17}$$

and for the t-statistic:

$$t_A^* = \lambda DF_\mu + \sqrt{1-\lambda^2}N(0,1), \tag{A.18}$$

where $[W_4(r), W_5(r)]'$ is the bivariate Brownian motion with correlation λ. Next, we will show that

$$\rho = \lambda. \tag{A.19}$$

Note that $\gamma = E(w_t w_t')^{-1}E(w_t e_t)$, so we have:

$$\sigma_v^2 = \sigma_e^2 - E(e_t w_t')E(w_t w_t')^{-1}E(w_t e_t). \tag{A.20}$$

Also, $E(w_t e_t) = C_{21} - \sigma_e^2 D_2$ and $E(w_t w_t') = C_{22} + \sigma_\varepsilon^2 D_2 D_2' - C_{21}D_2' - D_2 C_{21}'$. Therefore,

$$\sigma_v^2 = \sigma_e^2 - \left(C_{21} - \sigma_\varepsilon^2 D_2\right)'\left(C_{22} + \sigma_\varepsilon^2 D_2 D_2' - C_{21}D_2' - D_2 C_{21}'\right)^{-1}\left(C_{21} - \sigma_\varepsilon^2 D_2\right),$$

which is, from Lemma A1, $1/\sigma_\psi^2$. Therefore, $\rho = \lambda$.

Now we let $\hat{\beta}_A$ be the OLS estimator of β in the regression (18). The proof is complete if we show that $T\hat{\beta}_A$ and $T\hat{\beta}_A^*$ are identical asymptotically. Let $\hat{\zeta}_t = \left(\tilde{\xi}_t', \hat{w}_t'\right)'$, where $\tilde{\xi}_t = \xi_t - T^{-1}\sum_{t=1}^T \xi_t$. Then we have:

$$T\hat{\beta}_A = \frac{T^{-1}\left(\sum_{t=1}^T \tilde{y}_{t-1}e_t - \sum_{t=1}^T \tilde{y}_{t-1}\zeta_t'\left(\sum_{t=1}^T \tilde{\zeta}_t\tilde{\zeta}_t'\right)^{-1}\sum_{t=1}^T \tilde{\zeta}_t e_t\right)}{T^{-2}\left(\sum_{t=1}^T \tilde{y}_{t-1}^2 - \sum_{t=1}^T \tilde{y}_{t-1}\zeta_t'\left(\sum_{t=1}^T \tilde{\zeta}_t\tilde{\zeta}_t'\right)^{-1}\sum_{t=1}^T \zeta_t\tilde{y}_{t-1}\right)},$$

Since $T^{-1}\sum_{t=1}^T \hat{w}_t\xi_t' = o_p(1)$, and $T^{-1}\sum_{t=1}^T \tilde{\xi}_t e_t = o_p(1)$, we have:

$$T\hat{\beta}_A = \frac{T^{-1}\left(\sum_{t=1}^T \tilde{y}_{t-1}e_t - \sum_{t=1}^T \tilde{y}_{t-1}\hat{w}_t'\left(\sum_{t=1}^T \hat{w}_t\hat{w}_t'\right)^{-1}\sum_{t=1}^T \hat{w}_t' e_t\right)}{T^{-2}\left(\sum_{t=1}^T \tilde{y}_{t-1}^2\right)} + o_p(1).$$

Similarly,

$$T\hat{\beta}_A^* = \frac{T^{-1}\left(\sum_{t=1}^T \tilde{y}_{t-1}e_t - \sum_{t=1}^T \tilde{y}_{t-1}w_t'\left(\sum_{t=1}^T \tilde{w}_t\tilde{w}_t'\right)^{-1}\sum_{t=1}^T \tilde{w}_t'e_t\right)}{T^{-2}\left(\sum_{t=1}^T \tilde{y}_{t-1}^2\right)} + o_p(1),$$

$T\hat{\beta}_A$ and $T\hat{\beta}_A^*$ are asymptotically identical if $T^{-1}\sum \tilde{y}_{t-1}(\hat{w}_t - w_t) = o_p(1)$. However:

$$T^{-1}\sum \tilde{y}_{t-1}\hat{w}_t = T^{-1}\sum \tilde{y}_{t-1}\left[h(\varepsilon_t) + (\hat{\varepsilon}_t - \varepsilon_t)h'(\varepsilon_t) - \hat{\varepsilon}_t\hat{D}_2\right] + o_p(1)$$

Therefore,

$$\begin{aligned} & T^{-1}\sum \tilde{y}_{t-1}(\hat{w}_t - w_t) \\ = \; & T^{-1}\sum \tilde{y}_{t-1}\left[(\hat{\varepsilon}_t - \varepsilon_t)h'(\varepsilon_t) - (\hat{\varepsilon}_t - \varepsilon_t)\hat{D}_2 - \varepsilon_t\left(\hat{D}_2 - D_2\right)\right] + o_p(1) \end{aligned} \tag{A.21}$$

but,

$$T^{-1}\sum \tilde{y}_{t-1}(\hat{\varepsilon}_t - \varepsilon_t)h'(\varepsilon_t) = T\left(\hat{\beta} - \beta\right)T^{-2}\sum \tilde{y}_{t-1}^2 h'(\varepsilon_t) + o_p(1), \tag{A.22}$$

$$T^{-1}\sum \tilde{y}_{t-1}(\hat{\varepsilon}_t - \varepsilon_t)\hat{D}_2 = \hat{D}_2 T\left(\hat{\beta} - \beta\right)T^{-2}\sum \tilde{y}_{t-1}^2 + o_p(1), \tag{A.23}$$

and

$$T^{-1}\sum \tilde{y}_{t-1}\varepsilon_t\left(\hat{D}_2 - D_2\right) = o_p(1). \tag{A.24}$$

The two terms (A.22) and (A.23) cancel each other in the limit in (A.21), so the proof is complete.

References

[1] Amemiya, T. (1985), Advanced Econometrics, Basil Blackwell Ltd.

[2] Beelder, O. (1996), "Adaptive estimation and unit root tests," manuscript, Department of Economics, Rochester University.

[3] Cox, D.D. and I. Llatas (1991), "Maximum likelihood type estimation for nearly nonstationary autoregressive time series," *Annals of Statistics*, 19, 1109-1128.

[4] Hansen, B. E. (1995), "Rethinking the univariate approach to the unit root testing: using covariates to increase the power," *Econometric Theory*, 11, 1148-1171.

[5] Hasan, M.N. and R.W. Koenker (1997), "Robust rank test of unit root hypothesis," *Econometrica*, 65, 133-161.

[6] Herce, M.A. (1996), "Asymptotic theory of LAD estimation in a unit root process with finite variance errors," *Econometric Theory*, 12, 129-153.

[7] Im K.S. and P. Schmidt (2008), "More efficient estimation under non-normality when higher moments do not depend on the regressors, using residual-augmented least squares," *Journal of Econometrics*, 144, 219-233.

[8] Knight, K. (1989), "Limit theory for autoregressive-parameter estimates in an infinite-variance random walk," *The Canadian Journal of Statistics*, 17, 261-278.

[9] Lucas, A. (1995), "Unit root tests based on M-estimators," *Econometric Theory*, 11, 331-346.

[10] Shin, D. and B.S. So (1999), 'Unit Root Tests Based on Adaptive Maximum Likelihood Estimation," *Econometric Theory*, 15, 1-23.

Table 1
Rejection Ratio of Alternative Tests at $\alpha = 0.05$
No Serial Correlations

No Time Trend

Distributions		T=50 DF	T=50 AD	T=50 M5	T=50 RALS (2&3)	T=50 RALS (t5)	T=100 DF	T=100 AD	T=100 M5	T=100 RALS (2&3)	T=100 RALS (t5)
Normal	$\phi = 1$	.060	.043	.094	.054	.059	.051	.049	.069	.051	.051
	$\phi = 0.9$	.146	.091	.198	.132	.136	.352	.263	.346	.311	.330
Student t	$\phi = 1$	.058	.045	.052	.051	.050	.053	.067	.037	.051	.051
df=3	$\phi = 0.9$	.139	.197	.291	.270	.296	.358	.535	.649	.615	.676
Mixture	$\phi = 1$	.055	.040	.178	.044	.045	.058	.049	.130	.045	.044
Normal	$\phi = 0.9$	.145	.790	.217	.850	.916	.361	.991	.281	.995	.998
Chi-square	$\phi = 1$	.046	.048	.058	.043	.045	.052	.047	.036	.041	.045
df=1	$\phi = 0.9$	.126	.360	.332	.909	.339	.355	.796	.666	.999	.723
With Linear Trend											
Normal	$\phi = 1$	.062	.025	.148	.054	.057	.057	.035	.078	.054	.055
	$\phi = 0.9$	.109	.049	.204	.092	.099	.216	.129	.251	.191	.206
Student t	$\phi = 1$	.064	.026	.062	.060	.052	.054	.039	.036	.049	.049
df=3	$\phi = 0.9$	.100	.120	.231	.187	.192	.197	.386	.495	.441	.507
Mixture	$\phi = 1$	.054	.024	.292	.042	.044	.053	.027	.192	.042	.043
Normal	$\phi = 0.9$	.097	.628	.258	.669	.784	.219	.981	.255	.981	.996
Chi-square	$\phi = 1$	.055	.026	.064	.045	.051	.055	.038	.048	.039	.049
df=1	$\phi = 0.9$	.081	.251	.277	.797	.224	.202	.647	.506	.991	.529

AD and M5 denote the t-tests based on the adaptive MLE and the M-estimate assuming the error density is the student-t with 5 degres of freedom studied by Shin and So (1999) and Lucas (1995), respectively. Mixture normal is 0.5N(-3,1) + 0.5N(3,1). All the figures for AD and M5 have been reproduced from Shin and So (1999).

Table 2
Rejection Ratio of Alternative Tests at $\alpha = 0.05$
AR(1) error with AR coefficient 0.5, No Time Trend

T = 50

Distributions		ADF			RALS(2&3)			RALS(t5)		
		p=2	p=4	SC	p=2	p=4	SC	p=2	p=4	SC
Normal	$\phi = 1$	.056	.055	.071	.053	054	.061	.054	.053	.067
	$\phi = 0.9$	.100	.080	.116	.087	.074	.097	.093	.073	.110
Cauchy	$\phi = 1$	.075	.079	.055	.046	.053	.063	.048	.059	.051
	$\phi = 0.9$	.074	.074	.088	.664	.599	.698	.567	.504	.568
Student t	$\phi = 1$	.063	.060	.056	.050	.054	.064	.049	.056	.055
df=2	$\phi = 0.9$	.080	.069	.089	.300	.252	.326	.343	.281	.341
Double	$\phi = 1$	.051	.053	.059	.048	.050	.057	.049	.051	.058
Exponential	$\phi = 0.9$	.091	.084	.109	.131	.110	.150	.151	.120	.161
Chi-square	$\phi = 1$	.051	.058	.061	.050	.045	.032	.052	.051	.061
4 df	$\phi = 0.9$	.094	.080	.110	.260	.202	.191	.091	.081	.106
Beta(2,2)	$\phi = 1$	.060	.055	.073	.057	.050	.059	.053	.047	.060
	$\phi = 0.9$	.100	.087	.121	.126	.101	.133	.131	.103	.149

T = 100

		p=3	p=6	SC	p=3	p=6	SC	p=3	p=6	SC
Normal	$\phi = 1$	.055	.053	.061	.056	.048	.052	.055	.051	.056
	$\phi = 0.9$	.217	.163	.243	.196	.142	.217	.207	.145	.230
Cauchy	$\phi = 1$	.080	.076	.055	.040	.042	.055	.045	.050	.042
	$\phi = 0.9$	.144	.125	.181	.907	.852	.943	.796	.775	.803
Student t	$\phi = 1$	.053	.053	.050	.049	.052	.067	.050	.047	.049
df=2	$\phi = 0.9$	.190	.134	.220	.610	.512	.678	.716	.616	.740
Double	$\phi = 1$	.059	.053	.062	.055	.050	.063	.055	.047	.056
Exponential	$\phi = 0.9$	.216	.155	.246	.321	.231	.362	.377	.273	.400
Chi-square	$\phi = 1$	.052	.048	.052	.046	.048	.025	.050	.045	.047
df=4	$\phi = 0.9$	.224	.156	.242	.629	.480	.556	.217	.157	.237
Beta(2,2)	$\phi = 1$	.057	.053	.063	.048	.048	.052	.050	.046	.054
	$\phi = 0.9$	.216	.155	.246	.324	.225	.355	.343	.235	.376

Table 3
Rejection Ratio of Alternative Tests at $\alpha = 0.05$
AR(1) error with AR coefficient 0.5, with Linear Time Trend

T = 50

Distributions		ADF			RALS(2&3)			RALS(t5)		
		p=2	p=4	SC	p=2	p=4	SC	p=2	p=4	SC
Normal	$\phi = 1$	.053	.048	.074	.049	.044	.060	.050	.044	.070
	$\phi = 0.9$	.080	.069	.106	.069	.056	.081	.077	.059	.097
Cauchy	$\phi = 1$	.060	.058	.055	.064	.066	.078	.099	.103	.094
	$\phi = 0.9$	.073	.067	.067	.591	.516	.631	.370	.319	.383
Student t	$\phi = 1$	.057	.049	.060	.054	.060	.073	.058	.055	.063
df=2	$\phi = 0.9$	.078	.062	.081	.252	.201	.281	.248	.195	.255
Double	$\phi = 1$	.058	.053	.075	.055	.048	.074	.056	.049	.071
Exponential	$\phi = 0.9$	.083	.064	.098	.106	.085	.130	.115	.099	.135
Chi-square	$\phi = 1$	.058	.053	.073	.053	.049	.032	.053	.049	.065
df=4	$\phi = 0.9$	.074	.066	.094	.204	.133	.128	.074	.064	.091
Beta(2,2)	$\phi = 1$	.059	.050	.080	.050	.044	.058	.052	.044	.066
	$\phi = 0.9$	.082	.065	.111	.090	.072	.095	.092	.072	.112

T = 100

		p=3	p=6	SC	p=3	p=6	SC	p=3	p=6	SC
Normal	$\phi = 1$	.055	.053	.061	.053	.048	.057	.053	.053	.058
	$\phi = 0.9$	.155	.109	.175	.136	.094	.153	.139	.101	.167
Cauchy	$\phi = 1$	.061	.055	.031	.050	.057	.074	.120	.120	.108
	$\phi = 0.9$	.110	.100	.115	.867	.800	.925	.652	.627	.675
Student t	$\phi = 1$	.055	.045	.042	.049	.047	.071	.050	.050	.051
df=2	$\phi = 0.9$	.133	.090	.145	.525	.411	.608	.610	.495	.644
Double	$\phi = 1$	.059	.049	.061	.049	.045	.066	.051	.044	.059
Exponential	$\phi = 0.9$	.150	.105	.177	.223	.164	.276	.277	.196	.305
Chi-square	$\phi = 1$	.060	.049	.059	.050	.046	.021	.053	.044	.056
df=4	$\phi = 0.9$	.149	.102	.174	.524	.370	.414	.147	.100	.168
Beta(2,2)	$\phi = 1$	.052	.046	.064	.042	.036	.044	.043	.039	.047
	$\phi = 0.9$	.157	.106	.185	.235	.155	.259	.239	.159	.281

Table 4
Rejection Ratio of Alternative Tests at $\alpha = 0.05$
MA(1) error with MA coefficient –0.5, No Time Trend

T = 50

Distributions		ADF			RALS(2&3)			RALS(t5)		
		p=2	p=4	SC	p=2	p=4	SC	p=2	p=4	SC
Normal	$\phi = 1$	.088	.054	.095	.079	.051	.079	.085	.052	.091
	$\phi = 0.9$	.227	.109	.231	.190	.091	.182	.205	.097	.208
Cauchy	$\phi = 1$	.102	.077	.079	.180	.092	.194	.109	.068	.108
	$\phi = 0.9$	.163	.087	.177	.853	.712	.867	.625	.557	.622
Student t	$\phi = 1$	.091	.060	.079	.109	.060	.123	.102	.059	.101
df=2	$\phi = 0.9$	.198	.089	.199	.538	.347	.547	.553	.379	.539
Double	$\phi = 1$	.083	.053	.085	.088	.050	.094	.087	.054	.090
Exponential	$\phi = 0.9$	.219	.104	.216	.285	.149	.289	.320	.168	.313
Chi-square	$\phi = 1$	.085	.050	.088	.108	.054	.070	.079	.051	.082
df=4	$\phi = 0.9$	.223	.102	.227	.522	.282	.425	.209	.098	.212
Beta(2,2)	$\phi = 1$	.095	.054	.101	.086	.051	.084	.087	.050	.090
	$\phi = 0.9$	.237	.112	.241	.259	.135	.250	.276	.140	.277

T = 100

Distributions		ADF			RALS(2&3)			RALS(t5)		
		p=3	p=6	SC	p=3	p=6	SC	p=3	p=6	SC
Normal	$\phi = 1$	.050	.049	.057	.052	.046	.053	.050	.046	.056
	$\phi = 0.9$	.260	.187	.287	.230	.172	.257	.246	.176	.269
Cauchy	$\phi = 1$	.078	.073	.049	.041	.044	.057	.040	.042	.036
	$\phi = 0.9$	.171	.134	.207	.938	.889	.962	.785	.776	.796
Student t	$\phi = 1$	.053	.048	.048	.050	.047	.065	.048	.046	.050
df=2	$\phi = 0.9$	.235	.159	.266	.682	.570	.745	.771	.666	.798
Double	$\phi = 1$	.058	.051	.059	.052	.050	.058	.053	.048	.054
Exponential	$\phi = 0.9$	.263	.187	.303	.379	.280	.444	.440	.321	.486
Chi-square	$\phi = 1$	.047	.044	.050	.046	.045	.027	.041	.043	.045
df=4	$\phi = 0.9$	.254	.183	.293	.705	.543	.650	.248	.174	.286
Beta(2,2)	$\phi = 1$	.051	.047	.059	.047	.046	.053	.049	.045	.052
	$\phi = 0.9$	.253	.186	.296	.372	.265	.407	.391	.280	.438

Table 5
Rejection Ratio of Alternative Tests at $\alpha = 0.05$
MA(1) error with MA coefficient –0.5, with Linear Time Trend

T = 50

Distributions		ADF			RALS(2&3)			RALS(t5)		
		p=2	p=4	SC	p=2	p=4	SC	p=2	p=4	SC
Normal	$\phi = 1$	.111	.054	.127	.091	.046	.096	.096	.051	.108
	$\phi = 0.9$	.176	.076	.191	.145	.067	.148	.157	.072	.170
Cauchy	$\phi = 1$	.102	.061	.085	.262	.125	.270	.130	.094	.126
	$\phi = 0.9$	.138	.080	.126	.784	.608	.801	.423	.346	.424
Student t	$\phi = 1$	.101	.055	.099	.147	.076	.166	.115	.066	.120
df=2	$\phi = 0.9$	.158	.071	.157	.441	.263	.467	.401	.245	.397
Double	$\phi = 1$	.114	.059	.119	.116	.057	.124	.120	.061	.125
Exponential	$\phi = 0.9$	.177	.082	.184	.226	.109	.237	.250	.122	.248
Chi-square	$\phi = 1$	.111	.059	.118	.144	.059	.087	.102	.052	.108
df=4	$\phi = 0.9$	.168	.071	.171	.420	.183	.295	.153	.071	.154
Beta(2,2)	$\phi = 1$	.121	.055	.135	.101	.052	.099	.101	.051	.108
	$\phi = 0.9$	.181	.076	.199	.186	.086	.179	.201	.091	.211

T = 100

		p=3	p=6	SC	p=3	p=6	SC	p=3	p=6	SC
Normal	$\phi = 1$	.089	.052	.134	.080	.048	.122	.083	.051	.130
	$\phi = 0.9$	.260	.123	.396	.230	.112	.336	.243	.119	.359
Cauchy	$\phi = 1$	.081	.056	.067	.156	.075	.310	.134	.111	.144
	$\phi = 0.9$	.188	.111	.306	.934	.839	.972	.658	.620	.680
Student t	$\phi = 1$	.079	.044	.105	.098	.050	.195	.090	.052	.133
df=2	$\phi = 0.9$	.233	.106	.360	.684	.473	.823	.745	.556	.820
Double	$\phi = 1$	.093	.052	.135	.084	.048	.152	.086	.048	.136
Exponential	$\phi = 0.9$	.261	.125	.396	.371	.185	.533	.436	.228	.570
Chi-square	$\phi = 1$	.085	.049	.127	.095	.048	.091	.080	.049	.120
df=4	$\phi = 0.9$	.268	.128	.397	.708	.429	.725	.253	.121	.373
Beta(2,2)	$\phi = 1$	.084	.048	.136	.077	.040	.113	.078	.040	.125
	$\phi = 0.9$	.265	.135	.400	.362	.189	.483	.377	.192	.519