

1192

July 13, 1988

Asymptotic Distributions of the Unit Root Tests  
When the Process is Nearly Stationary

Abbreviated Title: Unit Root Tests

Sastry G. Pantula<sup>\*</sup>

Department of Statistics  
North Carolina State University  
Raleigh, NC 27695-8203 - USA

\* Partially supported by the National Science Foundation under grant NSF-DMS-8610127.

Key Words and Phrases: Random walk, white noise, unit root tests, limiting distributions, autoregressive and moving average processes.

AMS 1980 Subject Classifications: Primary 62M10.

Asymptotic Distributions of Unit Root Tests When  
the Process is Nearly Stationary

by

Sastry G. Pantula  
North Carolina State University  
Raleigh, NC 27695-8203

ABSTRACT

Several test criteria are available for testing the hypothesis that the autoregressive polynomial of an autoregressive moving average process has a single unit root. Schwert (1988), using a Monte Carlo study, investigates the performance of some of the available test criteria. He concludes that the actual levels of the test criteria considered in his study are far from the specified levels when the moving average polynomial also has a root close to one. This paper studies the asymptotic null distribution of the test statistics for testing  $\rho = 1$  in the model  $Y_t = \rho Y_{t-1} + e_t - \theta e_{t-1}$ , as  $\theta$  approaches one. It is shown that the test statistics differ from one another in their asymptotic properties depending on the rate at which  $\theta$  converges to one.

## 1. INTRODUCTION

Fuller (1976) and Dickey and Fuller (1979) have proposed simple tests for testing the hypothesis that the characteristic equation of an autoregressive process has a single unit root. The test criteria proposed by Dickey and Fuller (1979), however, are not valid when the process is an autoregressive moving average process. The moving average parameters in the model introduce some bias into the Dickey-Fuller test statistics, which are constructed assuming that the model is a pure autoregressive process. Said and Dickey (1984, 1985) propose two methods for testing for a single unit root in an autoregressive moving average process. Phillips (1987a) and Phillips and Perron (1988) suggest alternate criteria, which approximately correct the bias in the Dickey-Fuller test statistics. Hall (1988), on the other hand, suggests a criterion based on the instrumental variable approach which avoids the bias in the Dickey-Fuller criteria.

Schwert (1988) and Hall (1988) consider the model

$$Y_t = \rho Y_{t-1} + e_t - \theta e_{t-1}, \quad \rho = 1, \quad (1.1)$$

where  $e_t$  is a sequence of iid  $N(0,1)$  variables. Schwert (1988) cites several examples of economic data that satisfy model (1.1) with  $\theta$  close to one. See also Schwert (1987). Schwert (1988) considers the test statistics, for testing the null hypothesis  $\rho=1$ , proposed by Dickey and Fuller (1979), Said and Dickey (1984, 1985), Phillips (1987a) and Phillips and Perron (1988). Using a Monte Carlo study, Schwert (1988) compares the empirical level with the specified level of different test criteria. He concludes that the performance of the

test criteria is poor when  $\theta$  is close to one. When  $\theta = 0.8$ , the empirical levels of different test criteria are observed to be much higher than the specified level. Hall (1988) observes that his criteria based on the instrumental variable approach also reject the null hypothesis more often than expected, when  $\theta$  is close to unity.

The main goal of this paper is to study the asymptotic properties of the test criteria when  $\theta$  approaches unity. Note that the model (1.1) with  $\rho = 1$  and  $\theta = 1$  reduces to  $Y_t = e_t$ , a white noise process. The null hypothesis  $H_0: \rho=1, |\theta| < 1$ , includes processes that are arbitrarily close to a white noise process. If the process were a white noise process, one should reject the null hypothesis that the process is nonstationary ( $H_0: \rho=1, |\theta| < 1$ ). In finite samples, it may not be reasonable to expect that the above test criteria hold the proper level for all values of  $\theta$ , especially when  $\theta$  is arbitrarily close to one. To study the behavior of the test statistics when  $\theta$  is close to one, we consider the model (1.1) with  $\rho = 1$  and  $\theta = \theta_n = 1 - n^{-\delta}\gamma$ , where  $0 < \gamma < 2$  and  $\delta \geq 0$ . If  $\delta = 0$ , we get  $\theta = 1 - \gamma$  and the process is an integrated invertible moving average process. As  $\delta$  increases,  $\theta$  gets closer to one and the process behaves like a white noise process.

In section 2 we discuss the model and the unit root test statistics. In section 3 we present the asymptotic distributions of the test statistics for different values of  $\delta$ . We conclude with some remarks in section 4. The proofs of the main results are given in the mathematical appendix.

2. THE MODEL AND THE TEST STATISTICS

Consider the simple model

$$Y_t(n) = \rho Y_{t-1}(n) + e_t - \theta_n e_{t-1}, \quad t=1,2,\dots,n, \quad (2.1)$$

where  $\rho = 1$ ,  $\theta_n = 1 - n^{-\delta}\gamma$ ,  $0 < \gamma < 2$ ,  $\delta \geq 0$  and  $e_t$  is a sequence of iid random variables with mean zero, variance  $\sigma^2$  and bounded fourth moments. Without loss of generality, we assume throughout the paper that  $\sigma^2 = 1$ ,  $e_0 = 0$  and  $Y_0 = 0$ .

Note that,

$$\begin{aligned} Y_t(n) &= Y_{t-1}(n) + e_t - e_{t-1} + n^{-\delta}\gamma e_{t-1} \\ &= e_t + n^{-\delta}\gamma X_{t-1} \end{aligned} \quad (2.2)$$

where  $X_t = \sum_{i=1}^t e_i$  is a random walk process. So, the process  $Y_t(n)$  is a combination of a white noise process and a random walk process. If  $\delta = 0$ , then  $Y_t(n)$  is a nonstationary process given by  $Y_t = Y_{t-1} + e_t - (1-\gamma)e_{t-1}$ . As  $\delta$  increases to infinity,  $Y_t(n)$  behaves more and more like a white noise process. (In fact, if  $\delta > 0.5$  then  $\max_{1 \leq t \leq n} n^{-\delta} |X_{t-1}|$  converges to zero almost surely.) In this sense, the process is nearly stationary (in fact, nearly white noise). The model (2.1) is motivated by the examples of economic data cited in Schwert (1988).

Ahtola and Tiao (1984), Phillips (1987b) and Chan and Wei (1987) consider an analogous specification of the autoregressive coefficient. They consider processes that are nearly nonstationary, where  $\rho = \rho_n$  converges to one as  $n$  increases. The purpose of their studies is to investigate the asymptotic power

of the tests under a sequence of local alternatives. We, on the other hand, are considering a sequence of models in the null hypothesis that approach the alternative (stationary models) and study the asymptotic level of the tests.

Now, let us consider the behavior of the sample autocovariance function  $\hat{\gamma}(h)$  of the  $Y_t(n)$  process given in (2.1). Note that

$$\begin{aligned}\hat{\gamma}(h) &= n^{-1} \sum_{t=h+1}^n Y_t(n) Y_{t-h}(n) \\ &= n^{-1} \sum_{t=h+1}^n e_t e_{t-h} + n^{-2\delta+1} \gamma^2 n^{-2} \sum_{t=h+1}^n X_{t-1}^2 + o_p(n^{-\delta}).\end{aligned}$$

Therefore, if  $0 < \delta < 0.5$ ,

$$n^{2\delta-1} \hat{\gamma}(h) \xrightarrow{D} \gamma^2 \Gamma, \quad h=0,1,2, \dots$$

where  $\Gamma = \int_0^1 W^2(t) dt$  and  $W(t)$  is a standard Brownian motion. Therefore, for  $0 < \delta < 0.5$ ,

$$\hat{\rho}(h) = [\hat{\gamma}(0)]^{-1} \hat{\gamma}(h) \xrightarrow{P} 1.$$

(We use the notation " $\xrightarrow{D}$ " to indicate the convergence in distribution and " $\xrightarrow{P}$ " to denote the convergence in probability. We will also use the notation " $\Rightarrow$ " to signify weak convergence of the probability measures on the function space  $D[0,1]$ , where  $D[0,1]$  is the space of real valued functions on the interval  $[0,1]$  that are right continuous and have finite limits.) On the other hand, if  $\delta > 0.5$ , then

$$\hat{\gamma}(h) \xrightarrow{P} 0 \quad \text{for } h > 0$$

and

$$\hat{\gamma}(0) \xrightarrow{P} 1.$$

In fact, if  $\delta > 0.75$ , then from Corollary 6.3.5.1 of Fuller (1976) it follows that, for any fixed  $h$ ,

$$n^{1/2} [\hat{\rho}(1), \dots, \hat{\rho}(h)] \xrightarrow{D} N(\mathbf{0}, I_h)$$

where  $I_h$  denotes an identity matrix of size  $h$ . Therefore, when  $\delta > 0.75$ , an  $\alpha$ -level test criterion based on the Q-statistic suggested by Ljung and Box (1978) is expected to reject the hypothesis that the process is white noise only 100% of the time. That is, when  $\delta > 0.75$ , in practice one would tend to believe that the process is white noise. A limited simulation of the processes  $Y_t(n)$  in (2.1) with  $\delta > 0.75$ , indicates that the time plot and the sample correlogram resemble that of a white noise process. (Is it really so bad that we conclude the process is white noise when  $\delta > 0.75$ ?)

We now present some of the test statistics that exist in the literature for testing the null hypothesis  $H_0: \rho = 1, |\theta| < 1$  in the model (1.1) and present the asymptotic distributions under the null hypothesis (i.e.,  $\rho=1$  and  $\delta=0$  in (2.1)).

1.  $\hat{\tau}$  statistic suggested by Dickey and Fuller (1979):

The regression t-statistic for testing the coefficient of  $Y_{t-1}$  is zero in the regression of  $Y_t$  on  $Y_{t-1}$ ,

$$\hat{\tau} = s_1^{-1} \left( \sum_{t=2}^n Y_{t-1}^2 \right)^{1/2} (\hat{\rho} - 1), \quad (2.3)$$

where

$$s_1^2 = (n-2)^{-1} \sum_{t=2}^n (Y_t - \hat{\rho} Y_{t-1})^2,$$

and

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

Phillips (1987a) has shown that, under the model (2.1) with  $\rho = 1$  and  $\delta = 0$ ,

$$\hat{\tau} \xrightarrow{D} - \frac{(1-\gamma) + \gamma^2 \xi}{[1+(1-\gamma)^2]^{1/2} (\gamma^2 \Gamma)^{1/2}} , \quad (2.4)$$

where

$$\Gamma = \int_0^1 W^2(t) dt ,$$

$$\xi = \frac{1}{2} W^2(1) - \frac{1}{2} ,$$

and  $W(t)$  is a standard Brownian motion. Note that if, in addition,  $\gamma=1$

then  $Y_t = Y_{t-1} + e_t$  and

$$\hat{\tau} \xrightarrow{D} \Gamma^{-1/2} \xi . \quad (2.5)$$

Fuller (1976) gives the percentiles of the limiting distribution given in (2.5).

2.  $t_k$  statistic suggested by Said and Dickey (1984):

The regression t-statistic for testing the coefficient of  $Y_{t-1}$  is zero in the regression of  $u_t = Y_t - Y_{t-1}$  on  $Y_{t-1}, u_{t-1}, \dots, u_{t-k+1}$ , where  $k = o(n^{1/4})$ . That is,

$$t_k = (s_k^2 G_k^{-1})^{-1/2} \hat{\eta}_1 , \quad (2.6)$$

where  $\hat{\eta}_1$  is the first coordinate of  $\hat{\eta} = G_k^{-1} g_k$ ,



$$G_k = \sum_{t=k+1}^n L_{t-1} L'_{t-1} ,$$

$$L'_{t-1} = (Y_{t-1}, u_{t-1}, \dots, u_{t-k+1}) ,$$

$$g_k = \sum_{t=k+1}^n L_{t-1} u_t ,$$

$$s_k^2 = (n-2k)^{-1} \sum_{t=k+1}^n (u_t - L_{t-1} \hat{\eta})^2 ,$$

and  $G''$  is the (1,1)th element of  $G_k^{-1}$ . Said and Dickey (1984) have shown that under the model (2.1) with  $\rho=1$  and  $\delta=0$ ,

$$t_k \xrightarrow{\mathcal{D}} \Gamma^{-1/2} \xi .$$

3.  $t_{GN}$  statistic proposed by Said and Dickey (1985):

Consider the one-step Gauss Newton estimator  $(\tilde{\rho}, \tilde{\theta})'$  obtained by minimizing  $\sum_{t=1}^n [e_t(\rho, \theta)]^2$ , starting with an initial estimator  $(\rho^*, \theta^*)'$ , where  $\rho^* - 1 = o_p(n^{-3/4})$  and  $\theta^* - \theta = o_p(n^{-1/4})$ . (Said and Dickey (1985) suggest using  $\rho^* = 1$  and  $\theta^* =$  Durbin's estimator of  $\theta$ . In Result A.1 of the appendix, we show that it is not necessary to take  $\rho^* = 1$ . We show that it is sufficient to have  $\rho^* - 1 = o_p(n^{-3/4})$  and  $\theta^* - \theta = o_p(n^{-1/4})$ . In fact, we show that, as long as the initial estimators are properly chosen, the iterated Gauss Newton estimator may be used to test  $\rho=1$ , instead of the one-step Gauss Newton estimator with  $\rho^*=1$ .) Then, under the model (2.1) with  $\rho=1$  and  $\delta=0$ , the t-statistic

$$t_{GN} = (s_{e_1}^2)^{-1/2} (\tilde{\rho} - 1) \tag{2.7}$$

$$\xrightarrow{\mathcal{D}} \Gamma^{-1/2} \xi ,$$

where

$$s_e^2 = n^{-1} \sum_{t=1}^n [e_t(\tilde{\rho}, \tilde{\theta})]^2 ,$$

$c_1$  is the (1,1)th element of  $(\sum_{t=1}^n F_t F_t')^{-1}$  and  $F_t = F_t(\rho^*, \theta^*)$  is the 2x1 vector of partial derivatives of  $e_t(\rho, \theta)$  with respect to  $\rho$  and  $\theta$ .

4.  $Z_\alpha$  statistic proposed by Phillips (1987a):

Let  $\hat{\rho}$  denote the least squares estimator in the regression of  $Y_t$  on  $Y_{t-1}$ . Then, under the model (2.1) with  $\rho=1$  and  $\delta=0$ , the statistic

$$Z_\alpha = n(\hat{\rho}-1) - (2n^{-2} \sum_{t=2}^n Y_{t-1}^2)^{-1} (s_{n,k}^2 - s_u^2) , \quad (2.8)$$

$$\xrightarrow{D} \Gamma^{-1} \xi , \quad (2.9)$$

where

$$s_u^2 = n^{-1} \sum_{t=1}^n u_t^2 ,$$

$$s_{n,k}^2 = s_u^2 + 2n^{-1} \sum_{j=1}^k \sum_{t=j+1}^n u_t u_{t-j} ,$$

$$u_t = Y_t - Y_{t-1} ,$$

and  $k = o(n^{1/4})$ . The percentiles of the limiting distribution (2.9) of  $Z_\alpha$  are tabulated in Table 8.5.1 of Fuller (1976).

5.  $Z_{IV}$  statistic suggested by Hall (1988):

Let  $\hat{\rho}_{IV}$  denote the instrumental variable estimator of  $\rho$  in the regression of  $Y_t$  on  $Y_{t-1}$  where  $Y_{t-2}$  is used as an instrument. (Hall (1988) actually suggests a class of estimators where  $Y_{t-\lambda}$ ,  $\lambda > 1$ , is used as an instrument.) Then, under the model (2.1) with  $\rho=1$  and  $\delta=0$ , the statistic

$$Z_{IV} = n(\hat{\rho}_{IV} - 1) \tag{2.10}$$

$$\xrightarrow{D} \Gamma^{-1} \xi .$$

In the literature, there are several other test statistics (e.g.,  $t_{k,\mu}$  and  $t_{k,\tau}$  of Said and Dickey (1984),  $Z_t$ ,  $Z_{\alpha\mu}$ ,  $Z_{t\mu}$ ,  $Z_{\alpha\tau}$  and  $Z_{t\tau}$  of Phillips and Perron (1988) and the instrumental variable versions of Phillips and Perron (1988) test statistics proposed by Hall (1988)). The purpose of the paper is to give some indication of what happens to some of the test statistics when the process is nearly stationary. The ideas presented here can be routinely extended to other test statistics and hence will not be presented here.

### 3. MAIN RESULTS

For different values of  $\delta$  in model (2.1), we now present the asymptotic distribution of the five test statistics described in Section 2. Recall that under model (2.1),  $Y_t(n) = e_t + n^{-\delta}\gamma X_{t-1}$ , where  $X_t = \sum_{i=1}^t e_i$ . As noted earlier, if  $\delta > 0.5$ , then  $n^{-\delta} \max_{1 \leq t \leq n} |X_{t-1}|$  converges to zero almost surely. This indicates that the behavior of  $Y_t(n)$  (and hence of the statistics) may depend on whether  $\delta > 0.5$  or not.

Note also that the test statistics  $t_k$  and  $Z_\alpha$  depend on  $k$  which is of order  $n^{1/4}$ . In fact, the t-statistic  $t_k$  is based on approximating the integrated moving average processes in (1.1) by a  $k$ th order autoregressive process. From (2.1), we have

$$\begin{aligned}
 u_t &= Y_t(n) - Y_{t-1}(n) \\
 &= -(1-n^{-\delta}\gamma)e_{t-1} + e_t \\
 &= -\sum_{\lambda=1}^{k-1} (1-n^{-\delta}\gamma)^\lambda u_{t-\lambda} + e_t - (1-n^{-\delta}\gamma)^k e_{t-k}. \tag{3.1}
 \end{aligned}$$

The remainder term  $(1-n^{-\delta}\gamma)^k e_{t-k}$  in the autoregressive approximation may or may not converge to zero. Since  $k=O(n^{1/4})$ , the coefficient  $(1-n^{-\delta}\gamma)^k$  converges to zero or one depending upon  $0 < \delta < 0.25$  or  $\delta > 0.25$ . Therefore, in presenting the limiting distributions of the test statistics we consider three intervals:

(i)  $0 < \delta < 0.25$ , (ii)  $0.25 < \delta < 0.5$  and (iii)  $\delta > 0.5$ .

### 3.1 Results for the Case $0 < \delta < 0.25$ : "Nonstationary Region"

The asymptotic distributions of the five statistics are summarized in the following theorem.

**Theorem 3.1:** Consider the model (2.1) with  $\rho=1$  and  $0 < \delta < 0.25$ . Assume that the initial estimators  $\rho^*$  and  $\theta^*$  used in the computation of  $t_{GN}$  are such that  $\rho^* - 1 = o_p(n^{-1})$  and  $\theta^* - \theta_n = o_p(n^{-0.5-0.5\delta})$ . Then,

$$(a) \quad n^{-\delta} \hat{\tau} \xrightarrow{D} - [2\gamma^2 \Gamma]^{-1/2},$$

$$(b) \quad t_k \xrightarrow{D} \Gamma^{-1/2} \xi,$$

$$(c) \quad t_{GN} \xrightarrow{D} \Gamma^{-1/2} \xi,$$

$$(d) \quad Z_\alpha \xrightarrow{D} \Gamma^{-1} \xi,$$

$$(e) \quad Z_{IV} \xrightarrow{D} \Gamma^{-1} \xi,$$

where the test statistics,  $\Gamma$  and  $\xi$  are defined in Section 2.

Notice that the test statistic  $\hat{\tau}$  diverges to negative infinity. More importantly, note that the remaining statistics have the same asymptotic distributions as in the case  $\delta=0$ . That is, even if the process is nearly stationary, the test statistics have the asymptotic distributions as in the case when the process is nonstationary, provided the rate at which the moving average parameter  $\theta_n = 1-n^\delta\gamma$  converges to one is not faster than  $n^{1/4}$ . In this sense, the test criteria based on  $t_k$ ,  $t_{GN}$ ,  $Z_\alpha$  and  $Z_{IV}$  are robust and are expected to retain the proper level for  $0 < \delta < 0.25$ .

### 3.2 Results for the case $0.25 < \delta < 0.5$ : "Grey Zone"

The following theorem summarizes the limiting distributions of the test statistics given in Section 2 for processes satisfying (2.1) with  $\rho=1$  and  $0.25 < \delta < 0.5$ .

Theorem 3.2: Consider the model (2.1) with  $\rho=1$  and  $0.25 < \delta < 0.5$ . Then

$$(a) \quad n^{-\delta} \hat{\tau} \xrightarrow{\mathcal{D}} - [2\gamma^2\Gamma]^{-1/2} ,$$

$$(b) \quad n^{-\delta} t_k \xrightarrow{\mathcal{D}} -[\gamma^2\Gamma]^{-1/2} ,$$

$$(c) \quad n^{-2\delta+0.5} Z_\alpha \xrightarrow{\mathcal{D}} [\gamma^2\Gamma]^{-1} (N_3 - N_2) ,$$

$$(d) \quad n^{-2\delta+0.5} Z_{IV} \xrightarrow{\mathcal{D}} [\gamma^2\Gamma]^{-1} (N_1 - N_0) ,$$

where  $N_0$ ,  $N_1$ ,  $N_2$  and  $N_3$  are independent  $N(0,1)$  random variables independent of  $\Gamma$  and  $\xi$ .

Notice that the test statistics  $\hat{\tau}$  and  $t_k$  diverge to negative infinity.

The asymptotic distribution (properly normalized) of  $Z_\alpha$  is the same as that of  $Z_{IV}$ . Note that  $|Z_\alpha|$  and  $|Z_{IV}|$  diverge to infinity at the rate  $n^{2\delta-0.5}$ , whereas the test statistic  $t_k$  diverges to negative infinity at the rate  $n^{\delta-0.25}$ .

This indicates that the test statistic  $Z_\alpha$  and  $Z_{IV}$  are more unstable than the test statistic  $t_k$ . The asymptotic distribution of the test statistic  $t_{GN}$  is not clear and hence is not included.

### 3.3 Results for the case $\delta > 0.5$ : "Stationary Region"

The limiting distributions of  $\hat{\tau}$ ,  $t_k$ ,  $Z_\alpha$  and  $Z_{IV}$  for processes satisfying (2.1) with  $\delta > 0.5$  are summarized in the following theorem.

**Theorem 3.3:** Consider the model (2.1) with  $\rho = 1$  and  $\delta > 0.5$ . Then

$$(a) \quad n^{-1/2} \hat{\tau} \xrightarrow{P} -1 ,$$

$$(b) \quad n^{-\delta} t_k \xrightarrow{D} -[\gamma^2 \Gamma]^{-1/2} , \text{ if } 0.5 < \delta < 0.625$$

$$(c) \quad n^{-1/2} k^{1/2} t_k \xrightarrow{P} -1 , \text{ if } \delta > 0.625$$

$$(d) \quad n^{-1/2} Z_\alpha \xrightarrow{D} N_3 - N_2 ,$$

$$(e) \quad n^{-2\delta+0.5} Z_{IV} \xrightarrow{D} [\gamma^2 \Gamma]^{-1} [N_1 - N_0] , \text{ if } 0.5 < \delta < 0.75$$

and

$$(f) \quad n^{-1} Z_{IV} \xrightarrow{D} -1 + N_1^{-1} N_0 , \text{ if } \delta > 0.75 ,$$

where  $N_0$ ,  $N_1$ ,  $N_2$  and  $N_3$  are independent  $N(0,1)$  variables independent of  $\Gamma$  and  $\xi$ .

Note that  $\hat{\tau}$  and  $t_k$  diverge to negative infinity, whereas  $(2n)^{-1/2}Z_\alpha$  converges in distribution to a standard normal variable. Therefore, the probability that  $Z_\alpha$  less than a fixed critical value (negative) converges to 0.5. That is, in the long run, when  $\delta > 0.5$ , the criterion based on  $Z_\alpha$  rejects the hypothesis that the process is nonstationary 50% of the time. The test criterion based on  $t_k$ , on the other hand, almost always rejects the hypothesis that the process is nonstationary. Recall that when  $\delta > 0.5$ , the process  $Y_t(n)$  behaves more like a stationary process than like a nonstationary process and hence the more often the test statistic rejects the nonstationarity the better it is. In this sense, the t-statistic  $t_k$  performs better than the  $Z_\alpha$  statistic. However, if you interpret the probability of rejecting nonstationarity of the model (2.1) as the level of the test for nonstationarity of the model (1.1), then the statistic  $Z_\alpha$  is doing a "better" job than the statistic  $t_k$ .

The order in probability of  $Z_{IV}$  is greater than that of  $Z_\alpha$ . Also, the order in probability of  $Z_\alpha$  is greater than that of  $t_k$ , indicating that  $t_k$  is more stable than  $Z_\alpha$  and  $Z_{IV}$ . Note also that the estimator  $\hat{\rho}_{IV} = 1+n^{-1}Z_{IV}$  converges in distribution to a Cauchy random variable.

The behavior of the test statistic  $t_{GN}$  based on the Gauss Newton estimator is not easy to study under the framework of this paper. Recall that the iterated Gauss Newton estimator is the value of the vector  $(\rho, \theta)'$  that minimizes

$$Q_n(\rho, \theta) = n^{-1} \sum_{t=1}^n [e_t(\rho, \theta)]^2 ,$$

where

$$e_t(\rho, \theta) = Y_t + (\theta - \rho) \sum_{j=1}^{t-1} \theta^{j-1} Y_{t-j}.$$

Now, if  $Y_t = Y_t(n)$  satisfies (2.1) then, after some algebraic simplification, we can show that

$$e_t(\rho, \theta) = (\theta - \rho) Z_{t-1}(\theta) + e_t + n^{-\delta} \gamma [(1 - \theta)^{-1} (1 - \rho) X_{t-2} - (1 - \theta)^{-1} (\theta - \rho) Z_{t-2}(\theta) + e_{t-1}],$$

where  $Z_{t-1}(\theta) = \sum_{j=1}^{t-1} \theta^{j-1} e_{t-j}.$

Typically, in the iterative procedures an upperbound on  $\theta$  is set (e.g.,  $\theta < 0.97$  in Said and Dickey (1985)). That is, usually the search is limited to  $|\theta| < \lambda < 1$ . Therefore, if  $\delta > 0.5$  and  $|\theta| < \lambda < 1$ , then with probability one,

$$Q_n(\rho, \theta) \rightarrow Q_\infty(\rho, \theta) = (1 - \theta^2)^{-1} (\theta - \rho)^2 + 1,$$

uniformly in  $\rho$  and  $\theta$ . The function  $Q_\infty(\rho, \theta)$  is minimized whenever  $\theta = \rho$ . The matrix  $n^{-1} \sum_{t=1}^n F_t F_t'$  where  $F_t$  is the vector of partial derivatives of  $e_t(\rho, \theta)$  with respect to  $\rho$  and  $\theta$ , converges to a singular matrix. The asymptotic singularity of  $n^{-1} \sum_{t=1}^n F_t F_t'$  makes it difficult to study the properties of the t-statistic  $t_{GN}$ .

#### 4. SUMMARY AND REMARKS

Schwert (1988) cites several examples of economic data that are integrated first order moving average processes where the moving average parameter is close to one. Using an extensive Monte Carlo study, he compares the



performance of some of the existing test criteria for testing the hypothesis that the autoregressive parameter is equal to one when the moving average parameter is close to one. He observes that the empirical levels of the test criteria are far from the specified levels, even for large sample sizes. Similar results are also observed by Hall (1988). This paper provides the asymptotic distributions of some of the test statistics as the moving average parameter approaches one.

We have considered nearly stationary (white noise) processes given by  $Y_t(n) = Y_{t-1}(n) + e_n - (1-n^{-\delta}\gamma)e_{t-1}$ . A nearly stationary process is a combination of a white noise process and a random walk process. If  $\delta$  is small, the random walk part is dominant and  $Y_t(n)$  behaves like a nonstationary process. If  $\delta$  is large, the process behaves like a white noise process. (Note that if  $\theta_n = 1-n^{-\delta}\gamma$ , then  $\delta = -[\ln n]^{-1} [\ln(1-\theta_n) - \ln\gamma]$ . For example, with  $\gamma=1$ , the parameter  $\theta=0.8$  corresponds to  $\delta$  ranging from 0.5 to 0.26 as  $n$  changes from 25 to 500. The value  $\theta=0.8$  is considered by Schwert (1988) and Hall (1988).)

We have shown that the test statistics  $t_k$ ,  $t_{GN}$ ,  $Z_\alpha$  and  $Z_{IV}$  have the same asymptotic distributions for  $0 < \delta < 0.25$  as in the case  $\delta = 0$  (which corresponds to an invertible integrated moving average process). We have also shown that the test statistics are unstable when  $\delta > 0.25$ , that is, the order in probability of the statistics increases with the sample size  $n$ . The  $t$ -statistic  $t_k$  in particular diverges to negative infinity for  $\delta > 0.25$ . Recall that for  $\delta > 0.5$ , the process behaves like a white noise process and the criterion based on  $t_k$  rejects the hypothesis of nonstationarity almost always. The test criteria based on  $Z_\alpha$ , however, is very unstable and is expected to

reject the hypothesis of nonstationarity only 50% of the time when  $\delta > 0.5$ . No such conclusions about probability of rejection can be drawn for criteria based on  $Z_{IV}$  and  $t_{GN}$ . On the other hand the test statistics  $Z_\alpha$  and  $Z_{IV}$  have definite computational advantages over the test statistics  $t_k$  and  $t_{GN}$ . The matrices that need to be inverted in the computation of  $t_k$  and  $t_{GN}$  are converging to singular matrices asymptotically. This may cause some computational problems.

The test statistic  $\hat{\tau}$  is not appropriate for testing the hypothesis of nonstationarity when the process is a mixed model. We included  $\hat{\tau}$  in the paper only to indicate what happens to the test statistic  $\hat{\tau}$  when the model is incorrectly specified to be a first order autoregressive model.

The performance of the test statistics  $t_k$  and  $Z_\alpha$  depends on the number of lags  $k$  used in the computation. We have taken  $k = O(n^{1/4})$ . The orders of  $k$  and  $1-\theta_n$  both effect the asymptotic distributions of  $t_k$  and  $Z_\alpha$ . Consider for example  $\delta = 0.3 (> 0.25)$ . For  $\delta = 0.3$ , one may be able to show that the t-statistic  $t_k$  has the same asymptotic distribution as in the case of  $\delta=0$ , provided  $k$  is taken to be of order  $n^{0.31}$ . As we expect, when  $\theta$  is closer to one, it takes a larger number of lags to approximate an integrated moving average process by a higher order autoregressive process. In practice it is, however, difficult to decide what the value of  $\delta$  is and what multiplier of  $n^{1/4}$  one should use to select  $k$ .

As pointed out earlier, the composite hypothesis that the process has a single unit root (nonstationary) includes processes that are close to white noise. The present paper studies the asymptotic behavior of the different test statistics as the models in the null hypothesis approach stationary models.

Our paper concentrated on test statistics that did not include an intercept and/or a trend. For processes satisfying (2.1), with minor modifications, the results can be extended to test statistics where an intercept is (and a trend are) estimated. Also, the results for  $\delta = 0.25, 0.5, 0.625$  and  $0.75$  are not difficult. The extensions to higher order autoregressive and moving average processes, however, is not immediate and will be considered elsewhere.

MATHEMATICAL APPENDIX

Before we present the proofs for theorems in Section 3, we prove a result and make a few comments regarding the iterated Gauss Newton estimator. To keep the proof simple, we use the notation given in Chapter 8 of Fuller (1976) and Said and Dickey (1985).

Result A.1: Consider the model

$$Y_t = \rho Y_{t-1} + e_t + \beta e_{t-1}, \quad (A.1)$$

where  $\rho = 1$ ,  $|\beta| < 1$ ,  $e_0 = 0$ ,  $Y_0 = 0$  and  $\{e_t\}$  is a sequence of iid  $(0,1)$  random variables with bounded fourth moments. Let  $\rho^*$  and  $\beta^*$  be initial estimators of  $\rho$  and  $\beta$  such that  $\rho^* - 1 = o_p(n^{-3/4})$ ,  $|\beta^*| < 1$  and  $\beta^* - \beta = o_p(n^{-1/4})$ . (Note that in (A.1),  $\beta$  is the same as  $-\theta$  in the model (1.1).) Then, the one step Gauss Newton estimator  $(\tilde{\rho}, \tilde{\beta})'$  is such that

$$D_n(\tilde{\rho}-1, \tilde{\beta}-\beta)' \xrightarrow{D} (B_1^*, B_2^*)',$$

where

$$B_1^* = \Gamma^{-1} \xi \quad (A.2)$$

$$B_2^* \sim N(0, 1-\beta^2)$$

and

$$D_n = \text{diagonal}(n, n^{1/2}), \quad (A.3)$$

where  $\Gamma$  and  $\xi$  are defined in Section 2. Also, the t-statistic

$$t_{GN} = (s_e^2 c_1)^{-1/2} (\tilde{\rho}-1) \quad (A.4)$$

where

$$s_e^2 = n^{-1} \sum_{t=1}^n [e_t(\tilde{\rho}, \tilde{\beta})]^2 ,$$

$$c_1 = (1,1)\text{th element of } \left( \sum_{t=1}^n F_t^* F_t^{*'} \right)^{-1} ,$$

and  $F_t^* = F_t^*(\rho^*, \beta^*)$  is the 2x1 vector of partial derivatives of  $e_t(\rho, \beta)$  with respect to  $\rho$  and  $\beta$ .

Proof: Note that

$$\begin{aligned} e_t(\rho, \beta) &= Y_t - \sum_{j=1}^{t-1} (\rho + \beta)(-\beta)^{j-1} Y_{t-j} \\ &= Y_t - \rho Y_{t-1} - \beta e_{t-1}(\rho, \beta) . \end{aligned}$$

Define,

$$e_t(\rho^*, \beta^*) = Y_t - \rho^* Y_{t-1} - \beta^* e_{t-1}(\rho^*, \beta^*) , \quad t > 0$$

where  $e_0(\rho^*, \beta^*) = 0$  and  $(\rho^*, \beta^*)'$  is an initial estimator of  $(\rho, \beta)'$ . Expanding the function  $e_t(\rho, \beta)$  in a first order Taylor series about the point  $(\rho^*, \beta^*)'$ , we get

$$e_t(\rho, \beta) = e_t(\rho^*, \beta^*) - V_t^*(\rho - \rho^*) - W_t^*(\beta - \beta^*) + R_t ,$$

where

$$R_t = 0.5[\hat{S}_t(\rho^* - \rho)^2 + \hat{T}_t(\beta^* - \beta)^2 + 2\hat{K}_t(\rho^* - \rho)(\beta^* - \beta)],$$

$V_t^*$  and  $W_t^*$  are negative partial derivatives of  $e_t(\rho, \beta)$  with respect to  $\rho$  and  $\beta$  evaluated at  $(\rho^*, \beta^*)'$  and  $\hat{S}_t, \hat{T}_t$  and  $\hat{K}_t$  are second order partial derivatives of  $e_t(\rho, \beta)$  with respect to  $\rho, \beta$  and  $\rho$  and  $\beta$ , respectively, evaluated at a point on the line segment joining  $(\rho, \beta)'$  and  $(\rho^*, \beta^*)'$ .

From Said and Dickey (1985), we get for  $t > 0$ ,

$$V_t(\rho, \beta) = Y_{t-1} - \beta V_{t-1}(\rho, \beta) ,$$

$$W_t(\rho, \beta) = e_{t-1} - \beta W_{t-1}(\rho, \beta) ,$$

$$S_t(\rho, \beta) = \beta S_{t-1}(\rho, \beta)$$

$$T_t(\rho, \beta) = 2W_{t-1}(\rho, \beta) - \beta T_{t-1}(\rho, \beta)$$

and

$$K_t(\rho, \beta) = V_{t-1}(\rho, \beta) - \beta K_{t-1}(\rho, \beta)$$

where  $V_0 = 0, W_0 = 0, S_0 = 0, T_0 = 0$  and  $K_0 = 0$ . Note that  $S_t \equiv 0$ .

Let  $\hat{\delta}_1$  and  $\hat{\delta}_2$  denote the regression coefficients in the regression of  $e_t(\rho^*, \beta^*)$  on  $V_t^*$  and  $W_t^*$ . That is,

$$\begin{bmatrix} \hat{\delta}_1 \\ \hat{\delta}_2 \end{bmatrix} = \begin{bmatrix} n & & n \\ \Sigma & V_t^{*2} & \Sigma V_t^* W_t^* \\ t=1 & & t=1 \end{bmatrix}^{-1} \begin{bmatrix} n \\ \Sigma V_t^* e_t(\rho^*, \beta^*) \\ t=1 \\ n \\ \Sigma W_t^* e_t(\rho^*, \beta^*) \\ t=1 \end{bmatrix} .$$

The one-step Gauss Newton estimator  $(\tilde{\rho}, \tilde{\beta})'$  is given by

$$(\tilde{\rho}, \tilde{\beta})' = (\rho^*, \beta^*)' + (\hat{\delta}_1, \hat{\delta}_2)' .$$

Therefore,

$$D_n(\tilde{\rho}-1, \tilde{\beta}-\beta)' = \begin{bmatrix} n^{-2} \sum_{t=1}^n V_t^{*2} & n^{-3/2} \sum_{t=1}^n V_t^* W_t^* \\ n^{-3/2} \sum_{t=1}^n W_t^* V_t^* & n^{-1} \sum_{t=1}^n W_t^{*2} \end{bmatrix}^{-1} \begin{bmatrix} n^{-1} \sum_{t=1}^n V_t^* (e_t - R_t) \\ n^{-1/2} \sum_{t=1}^n W_t^* (e_t - R_t) \end{bmatrix} ,$$

where  $D_n = \text{diagonal}(n, n^{1/2})$ . Note that  $V_t^*$  is a function of  $\beta^*$  alone and hence from Said and Dickey (1985) we have

$$\begin{aligned} (1+\beta^*)V_t^* &= (1+\beta)X_{t-1} + f_t , \\ n^{-2} \sum_{t=1}^n V_t^{*2} &= n^{-2} \sum_{t=1}^n X_{t-1}^2 + o_p(n^{-1/4}) , \\ n^{-1} \sum_{t=1}^n V_t^* e_t &= n^{-1} \sum_{t=1}^n X_{t-1} e_t + o_p(n^{-1/4}) , \end{aligned}$$

where

$$X_t = \sum_{i=1}^t e_t$$

and

$$f_t = \beta^* (V_t^* - V_{t-1}^*) - \beta e_{t-1} .$$

Let  $u_t = Y_t - Y_{t-1} = e_t + \beta e_{t-1}$ . Then,

$$e_t(\rho^*, \beta^*) + \beta^* e_{t-1}(\rho^*, \beta^*) = u_t + (1-\rho^*)Y_{t-1}$$

and hence,

$$e_t(\rho^*, \beta^*) = \sum_{j=0}^{t-1} (-\beta^*)^j [u_{t-j} + (1-\rho^*)Y_{t-1-j}] .$$

After some algebraic manipulations, it can be shown that

$$(1+\beta^*)e_t(\rho^*, \beta^*) = (1+\beta)(1-\rho^*)X_{t-1} + (\beta^* + \rho^*)(\beta-\beta^*) \sum_{j=1}^{t-1} (-\beta^*)^{j-1} e_{t-j} .$$

Also, it can be shown that

$$\begin{aligned} W_t^* &= \sum_{j=0}^{t-2} (-\beta^*)^j e_{t-1-j}(\rho^*, \beta^*) \\ &= W_t(1, \beta^*) + (1+\beta^*)^{-2}(1-\rho^*)X_{t-2} + (1-\rho^*)(1+\beta^*)^{-1}A_{t-2}^* , \end{aligned}$$

where

$$A_{t-2}^* = \sum_{j=1}^{t-2} (-\beta^*)^{j-1} [(1+\beta^*)^{-1}\beta^* + (\beta^* - \beta)j] e_{t-1-j}$$

and

$$W_t(1, \beta^*) = e_{t-1} + \sum_{j=1}^{t-2} (-\beta^*)^{j-1} [(\beta-\beta^*)j - \beta^*] e_{t-1-j} .$$

Note that  $W_t(1, \beta^*)$  is the same as  $\hat{W}_t$  considered by Said and Dickey (1985) when the initial estimator for  $\rho$  is taken to be one.

Now, it is easy to show that

$$n^{-1} \sum_{t=1}^n W_t^{*2} = n^{-1} \sum_{t=1}^n [W_t(1, \beta^*)]^2 + o_p(n^{-1/2}) ,$$

$$n^{-3/2} \sum_{t=1}^n V_t^* W_t^* = o_p(1) ,$$

and



$$n^{-1/2} \sum_{t=1}^n W_t^* e_t = n^{-1/2} \sum_{t=1}^n W_t(1, \beta^*) e_t + o_p(1) .$$

To complete the proof, we need to show that  $n^{-1} \sum_{t=1}^n V_t^* R_t$  and  $n^{-1/2} \sum_{t=1}^n W_t^* R_t$  converge to zero in probability. Note that

$$\begin{aligned} K_t(\rho, \beta) &= (-\beta)K_{t-1}(\rho, \beta) + V_{t-1}(\rho, \beta) \\ &= \sum_{j=0}^{t-1} (-\beta)^j V_{t-1-j}(\rho, \beta) \\ &= \sum_{j=0}^{t-1} a_j e_{t-1-j} + a X_{t-1} \end{aligned}$$

and

$$\begin{aligned} T_t(\rho, \beta) &= (-\beta)T_{t-1}(\rho, \beta) + 2W_{t-1}(\rho, \beta) \\ &= 2 \sum_{j=0}^{t-1} (-\beta)^j W_{t-1-j}(\rho, \beta) \\ &= \sum_{j=0}^{t-1} b_j e_{t-1-j} , \end{aligned}$$

where  $a = o_p(1)$ ,  $|a_j| < M \lambda^j$  and  $|b_j| < M \lambda^j$  for some finite  $M$  and  $0 < \lambda < 1$ . Since  $\rho^* - 1 = o_p(n^{-3/4})$  and  $\beta^* - \beta = o_p(n^{-1/4})$ , we get

$$\begin{aligned} 2(1+\beta^*) n^{-1} \sum_{t=1}^n V_t^* R_t &= (1+\beta)(\beta^* - \beta)^2 n^{-1} \sum_{t=1}^n X_{t-1} \sum_{j=0}^{t-1} b_j e_{t-1-j} \\ &\quad + 2(1+\beta)(\rho^* - 1)(\beta^* - \beta) a n^{-1} \sum_{t=1}^n X_{t-1}^2 + o_p(1) \\ &= o_p(1) , \end{aligned}$$

and

$$\begin{aligned} 2n^{-1/2} \sum_{t=1}^n W_t^* R_t &= n^{-1/2} (\rho^* - 1) (\beta^* - \beta) \sum_{t=1}^n W_t (1, \beta^*) \hat{K}_t + o_p(1) \\ &= o_p(1) . \end{aligned}$$

Therefore, from Theorem 3.1 of Said and Dickey (1985) we get the desired result. **I**

The result (A.1) has several implications. First, it indicates that the one-step Gauss Newton estimator  $\tilde{\rho}$ , starting with any initial estimator  $\rho^*$  (not necessarily equal to one), can be used for testing the null hypothesis that  $\rho = 1$ , provided  $\rho^* - 1 = o_p(n^{-3/4})$ . Also, since  $\tilde{\rho} - 1 = o_p(n^{-1})$  and  $\tilde{\beta} - \beta = o_p(n^{-1/2})$ , we do not have to restrict to the one-step Gauss Newton estimator. That is, one can use the iterated Gauss Newton estimator for  $\rho$  to test the hypothesis that  $\rho = 1$ . The choice of the initial estimator is still important in the sense we must have  $\rho^* - 1 = o_p(n^{-3/4})$ . Said and Dickey (1985) suggest using  $\rho^* = 1$  and  $\beta^*$  to be the Durbin estimator

$$\beta^* = \begin{bmatrix} k-1 & & \\ \Sigma & \hat{c}_{i-1}^2 & \\ i=1 & & \end{bmatrix}^{-1} \begin{bmatrix} k-1 & & \\ \Sigma & \hat{c}_i & \hat{c}_{i-1} \\ i=1 & & \end{bmatrix}$$

where  $\hat{c}_i = \hat{\eta}_{i+1}$ ,  $i=1, \dots, k-1$ ;  $\hat{c}_0 = \hat{\eta}_1 + 1$ ;  $k = o(n^{1/4})$  and  $\hat{\eta} = (\hat{\eta}_1, \dots, \hat{\eta}_k)'$  are the least squares coefficients in the regression of  $u_t = Y_t - Y_{t-1}$  on  $Y_{t-1}, u_{t-1}, \dots, u_{t-k+1}$ . From the results of Said and Dickey (1985) it can be shown that

$$\begin{aligned} \hat{\rho}^* &= 1 + (1 + \beta^*) \hat{\eta}_1 \\ &= 1 + o_p(n^{-1}) \end{aligned}$$

and

$$\beta^* = \beta + o_p(n^{-1/4}) .$$

So, an alternative initial estimator for  $\rho$  is  $\hat{\rho}^*$ .

We will now present simple expressions for the test statistics, considered in Section 2, that will be useful in deriving the asymptotic distributions.

1. Consider the  $\hat{\tau}$  statistic proposed by Dickey and Fuller (1979),

$$\hat{\tau} = s_1^{-1} \left( \sum_{t=2}^n Y_{t-1}^2 \right)^{1/2} (\hat{\rho} - 1)$$

where

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

$$\begin{aligned} s_1^2 &= n^{-1} \sum_{t=2}^n (Y_t - \hat{\rho} Y_{t-1})^2 \\ &= n^{-1} \sum_{t=2}^n u_t^2 - (\hat{\rho} - 1)^2 n^{-1} \sum_{t=2}^n Y_{t-1}^2, \end{aligned} \tag{A.5}$$

and

$$u_t = Y_t - Y_{t-1} .$$

It is easy to see that

$$\hat{\tau} = \left( s_1^2 \sum_{t=2}^n Y_{t-1}^2 \right)^{-1/2} \sum_{t=2}^n Y_{t-1} u_t . \tag{A.6}$$

2. Consider the regression t-statistic  $t_k$  for testing the coefficient of  $Y_{t-1}$  is zero in the regression of  $u_t$  on  $Y_{t-1}, u_{t-1}, \dots, u_{t-k+1}$ , where  $k = O(n^{1/4})$ .

From (3.1), we have

$$u_t = - \sum_{\ell=1}^{k-1} (1-n^{-\delta}\gamma)^\ell u_{t-\ell} + e_t - (1-n^{-\delta}\gamma)^k e_{t-k} . \quad (\text{A.7})$$

Define, for  $j=1, \dots, k-1$ , and  $t > k$ ,

$$w_{t-j} = e_{t-j} - (1-n^{-\delta}\gamma)^{k-j} e_{t-k} \quad (\text{A.8})$$

$$= \sum_{i=0}^{k-j-1} (1-n^{-\delta}\gamma)^i u_{t-j-i} .$$

Since  $w_{t-1}, \dots, w_{t-k+1}$  are linear functions of  $u_{t-1}, \dots, u_{t-k+1}$ , the test statistic  $t_k$  is the same as the t-statistic for testing the coefficient of  $Y_{t-1}$  is zero in the regression of  $u_t$  on  $Y_{t-1}, w_{t-1}, \dots, w_{t-k+1}$ . Note that from (A.7) and (A.8) we have

$$u_t = -(1-n^{-\delta}\gamma)w_{t-1} + e_t - (1-n^{-\delta}\gamma)^k e_{t-k} . \quad (\text{A.9})$$

Let  $\hat{\eta}^*$  denote the regression least squares coefficients in the regression of  $u_t$  on  $Y_{t-1}, w_{t-1}, \dots, w_{t-k+1}$ . Then,

$$\hat{\eta}^* = H_k^{-1} h_k \quad (\text{A.10})$$

where

$$H_k = \begin{bmatrix} \sum_{t=k+1}^n Y_{t-1}^2 & \mathbf{b}'_k \\ \mathbf{b}_k & \mathbf{A}_k \end{bmatrix},$$

$$\mathbf{h}_k = \begin{bmatrix} \sum_{t=k+1}^n Y_{t-1} u_t \\ \mathbf{c}_k \end{bmatrix},$$

$$\mathbf{b}_k = \sum_{t=k+1}^n Y_{t-1} \mathbf{v}_{t-1},$$

$$\mathbf{c}_k = \sum_{t=k+1}^n u_t \mathbf{v}_{t-1},$$

$$\mathbf{v}_{t-1} = (w_{t-1}, \dots, w_{t-k+1})',$$

and

$$\mathbf{A}_k = \sum_{t=k+1}^n \mathbf{v}_{t-1} \mathbf{v}'_{t-1}.$$

Note that

$$H_k^{-1} = \begin{bmatrix} d_k & -d_k \mathbf{b}'_k \mathbf{A}_k^{-1} \\ -d_k \mathbf{A}_k^{-1} \mathbf{b}_k & \mathbf{A}_k^{-1} + \mathbf{A}_k^{-1} \mathbf{b}_k \mathbf{b}'_k \mathbf{A}_k^{-1} d_k \end{bmatrix}$$

where

$$d_k = \left[ \sum_{t=k+1}^n Y_{t-1}^2 - \mathbf{b}'_k \mathbf{A}_k^{-1} \mathbf{b}_k \right]^{-1}.$$

Therefore, the t-statistic

$$t_k = s_k^{-1} d_k^{1/2} \left[ \sum_{t=k+1}^n Y_{t-1} u_t - \mathbf{b}'_k \mathbf{A}_k^{-1} \mathbf{c}_k \right], \quad (\text{A.11})$$

where

$$s_k^2 = (n-2k)^{-1} \left[ \sum_{t=k+1}^n u_t^2 - \mathbf{c}'_k \mathbf{A}_k^{-1} \mathbf{c}_k - d_k \left\{ \sum_{t=k+1}^n Y_{t-1} u_t - \mathbf{b}'_k \mathbf{A}_k^{-1} \mathbf{c}_k \right\}^2 \right].$$

3. The test statistic  $t_{GN}$  is as given in (2.7).
4. It is easy to show that the test statistic  $Z_\alpha$  given in (2.8) can also be written as

$$Z_\alpha = (n^{-1} \sum_{t=1}^n Y_{t-1}^2)^{-1} \left[ \sum_{t=2}^n Y_{t-1} u_t - \sum_{j=1}^k \sum_{t=j+1}^n u_t u_{t-j} \right]. \quad (\text{A.12})$$

5. Finally, the test statistic  $Z_{IV}$  given in (2.10) can be written as

$$\begin{aligned} Z_{IV} &= n \left[ \left( \sum_{t=2}^n Y_{t-2} Y_{t-1} \right)^{-1} \left( \sum_{t=2}^n Y_{t-2} Y_t \right) - 1 \right] \\ &= n \left[ \left( \sum_{t=2}^n Y_{t-2} Y_{t-1} \right)^{-1} \sum_{t=2}^n Y_{t-2} u_t \right]. \end{aligned} \quad (\text{A.13})$$

We will now prove a Lemma that will be used in deriving the asymptotic distributions of the test statistics.

Lemma A.1: Consider the vector  $\alpha = (\alpha_1, \dots, \alpha_{k-1})'$  where  $\alpha_i = (1-n^{-\delta}\gamma)^{k-i}$ .

Let  $\mathbf{1} = (1, \dots, 1)'$  is a  $(k-1) \times 1$  vector with all coordinates equal to one.

Then, for  $k = O(n^{1/4})$ , as  $n$  tends to infinity,

$$(i) (1 - n^{-\delta}\gamma)^k \longrightarrow 0 \text{ if } 0 < \delta < 0.25 ,$$

$$(ii) 1 - (1 - n^{-\delta}\gamma)^k = n^{-\delta}\gamma k + O(n^{-2\delta}k^2) \text{ for } \delta > 0.25 ,$$

$$(iii) n^{-\delta}\alpha'\alpha \longrightarrow (2\gamma)^{-1} \text{ if } 0 < \delta < 0.25 ,$$

$$(iv) k^{-1}\alpha'\alpha = 1 + O(n^{-\delta}k) \text{ if } \delta > 0.25 ,$$

$$(v) n^{-\delta}1'\alpha \longrightarrow \gamma^{-1} \text{ if } 0 < \delta < 0.25 ,$$

$$(vi) k^{-1}1'\alpha = 1 + O(n^{-\delta}k) \text{ if } \delta > 0.25 ,$$

and

$$(viii) (1 - \alpha)'(1 - \alpha) = \begin{cases} O(k) & \text{if } 0 < \delta < 0.25 \\ O(n^{-\delta}k^2) & \text{if } \delta > 0.25 \end{cases} .$$

Proof: Note that

$$(1 - n^{-\delta}\gamma)^{n^{\delta}} \longrightarrow e^{-\gamma}$$

and

$$(1 - n^{-\delta}\gamma)^k = (1 - n^{-\delta}\gamma)^{n^{\delta}} kn^{-\delta} .$$

Therefore, for  $0 < \delta < 0.25$ ,  $(1 - n^{-\delta}\gamma)^k$  converges to zero (exponentially).

Result (ii) follows from the first order Taylor series expansion of  $(1 - n^{-\delta}\gamma)^k$ .

Now,

$$\begin{aligned}
 \alpha' \alpha &= \sum_{i=1}^{k-1} (1-n^{-\delta} \gamma)^{2i} \\
 &= \left[ 1 - (1-n^{-\delta} \gamma)^2 \right]^{-1} (1-n^{-\delta} \gamma)^2 \left[ 1 - (1-n^{-\delta} \gamma)^{2(k-1)} \right] \\
 &= \left[ 2\gamma - n^{-\delta} \gamma^2 \right]^{-1} (1-n^{-\delta} \gamma)^2 n^{\delta} \left[ 1 - (1-n^{-\delta} \gamma)^{2(k-1)} \right].
 \end{aligned}$$

Therefore, the results (iii) and (iv) follow from the results (i) and (ii). Also,

$$\begin{aligned}
 \mathbf{1}' \alpha &= \sum_{i=1}^{k-1} (1-n^{-\delta} \gamma)^i \\
 &= \gamma^{-1} n^{\delta} (1-n^{-\delta} \gamma) \left[ 1 - (1-n^{-\delta} \gamma)^{k-1} \right]
 \end{aligned}$$

and the results (v) and (vi) follow from the results (iii) - (iv). Finally,

$$(\mathbf{1} - \alpha)'(\mathbf{1} - \alpha) = k - 2\mathbf{1}'\alpha + \alpha'\alpha$$

and the result (vii) follows from the results (iii) - (vi). **■**

Notation: We use the Euclidean norm  $\|\mathbf{x}\| = (\mathbf{x}'\mathbf{x})^{1/2}$ , of a column vector  $\mathbf{x}$  to define a matrix norm  $\|\mathbf{B}\|$ , when

$$\|\mathbf{B}\| = \sup\{\|\mathbf{B}\mathbf{x}\| : \|\mathbf{x}\| < 1\}.$$

Recall that  $\|\mathbf{B}\|^2$  is bounded by the sum of squares of the elements of  $\mathbf{B}$  and that  $\|\mathbf{B}\|$  is bounded by the largest modulus of the eigenvalues of  $\mathbf{B}$ .



We are now ready to derive the asymptotic distribution of the test statistics. To save space, we will suppress  $n$  and write  $Y_t$  for  $Y_t(n)$ .

Proof of Theorem 3.1: Recall that

$$Y_t = e_t + n^{-\delta} \gamma X_{t-1} ,$$

$$X_t = \sum_{i=1}^t e_i$$

and we assume that  $0 < \delta < 0.25$  and  $0 < \gamma < 2$ .

(a) Note that

$$\sum_{t=2}^n Y_{t-1}^2 = \sum_{t=1}^{n-1} e_t^2 + n^{-2\delta} \gamma^2 \sum_{t=1}^{n-1} X_{t-1}^2 + 2n^{-\delta} \gamma \sum_{t=1}^{n-1} X_{t-1} e_t , \quad (\text{A.14})$$

and with  $u_t = Y_t - Y_{t-1} = e_t - (1-n^{-\delta} \gamma) e_{t-1}$ ,

$$\begin{aligned} \sum_{t=2}^n Y_{t-1} u_t &= -(1-n^{-\delta} \gamma) \sum_{t=2}^n e_{t-1}^2 + n^{-2\delta} \gamma^2 \sum_{t=2}^n X_{t-1} e_t \\ &\quad + (1-n^{-\delta} \gamma) \sum_{t=2}^n e_{t-1} e_t + n^{-\delta} \gamma X_{n-1} e_n . \end{aligned} \quad (\text{A.15})$$

Therefore,

$$n^{-1} \sum_{t=2}^n Y_{t-1} u_t \xrightarrow{P} -1$$

and for  $0 < \delta < 0.25$ ,

$$n^{-2+2\delta} \sum_{t=2}^n Y_{t-1}^2 \xrightarrow{D} \gamma^2 \Gamma .$$

Also, note that

$$n^{-1} \sum_{t=2}^n u_t^2 = n^{-1} \sum_{t=2}^n e_t^2 + (1-n^{-\delta}\gamma)^2 \sum_{t=2}^n e_{t-1}^2 - 2(1-n^{-\delta}\gamma)n^{-1} \sum_{t=2}^n e_{t-1}e_t$$

$$\xrightarrow{P} 2 . \quad (A.16)$$

Therefore, from (A.6),

$$n^{-\delta} \hat{\tau} = (s_1^2 n^{-2+2\delta} \sum_{t=2}^n Y_{t-1}^2)^{-1/2} n^{-1} \sum_{t=2}^n Y_{t-1} u_t$$

$$\xrightarrow{D} - [2\gamma^2 \Gamma]^{-1/2} .$$

(b) Consider the t-statistic  $t_k$  given in (A.11). Note that, from (A.15),

$$n^{-1} \sum_{t=k+1}^n Y_{t-1} u_t = -(1-n^{-\delta}\gamma) + n^{-2\delta}\gamma^2 n^{-1} \sum_{t=2}^n X_{t-1} e_t + o_p(n^{-1/2}). \quad (A.17)$$

Also, for  $i, j = 1, \dots, k-1$ ,

$$n^{-1} \sum_{t=k+1}^n w_{t-i} w_{t-j} = n^{-1} \sum_{t=k+1}^n [e_{t-i} - (1-n^{-\delta}\gamma)^{k-i} e_{t-k}] [e_{t-j} - (1-n^{-\delta}\gamma)^{k-j} e_{t-k}]$$

$$= \delta_{ij} + (1-n^{-\delta}\gamma)^{2k-i-j} + o_p(n^{-1/2}) ,$$

where  $\delta_{ij}$  is one if  $i=j$  and zero otherwise. Therefore,

$$Q = \|n^{-1} A_k - I - \alpha \alpha'\| = o_p(n^{-1/4}) ,$$

where  $A_k$  is defined following (A.10) and  $\alpha$  is defined in Lemma A.1. It is easy to show that the eigenvalues of  $(I + \alpha \alpha')^{-1}$  are  $(1 + \alpha' \alpha)^{-1}$  and one. Also,

$$(I + \alpha \alpha')^{-1} = I - (1 + \alpha' \alpha)^{-1} \alpha \alpha' . \quad (A.18)$$

Therefore,

$$p = \| (I + \alpha \alpha')^{-1} \| < 1$$

and

$$\begin{aligned} q &= \| (n^{-1} A_k)^{-1} - (I + \alpha \alpha')^{-1} \| \\ &\leq \| (n^{-1} A_k)^{-1} \| \| (n^{-1} A_k) - (I + \alpha \alpha') \| \| (I + \alpha \alpha')^{-1} \| \\ &\leq (q + p) Q P . \end{aligned}$$

Since  $Q = O_p(n^{-1/4})$  and  $p = O(1)$ , we get

$$q = O_p(n^{-1/4}) , \tag{A.19}$$

and

$$\| (n^{-1} A_k)^{-1} \| = O_p(1) . \tag{A.20}$$

Consider now, for  $j=1, \dots, k-1$ ,

$$\begin{aligned} n^{-1} \sum_{t=k+1}^n w_{t-j} u_t &= n^{-1} \sum_{t=k+1}^n [e_{t-j} - (1-n^{-\delta\gamma})^{k-j} e_{t-k}] [e_t - (1-n^{-\delta\gamma}) e_{t-1}] \\ &= -(1-n^{-\delta\gamma}) n^{-1} \sum_{t=k+1}^n e_{t-j} e_{t-1} + O_p(n^{-1/2}) . \end{aligned}$$

Therefore,

$$\| n^{-1} c_k + (1-n^{-\delta\gamma}) \epsilon_1 \| = O_p(k^{1/2} n^{-1/2}) , \tag{A.21}$$

where  $\epsilon_1 = (1, 0, \dots, 0)'$  and  $c_k$  is defined following (A.10). Also, for  $j=1, \dots, k-1$ ,

$$\begin{aligned} n^{-1} \sum_{t=k+1}^n w_{t-j} Y_{t-1} &= n^{-1} \sum_{t=k+1}^n \left[ e_{t-j} - (1-n^{-\delta}\gamma)^{k-j} e_{t-k} \right] \left[ e_{t-1} + n^{-\delta}\gamma X_{t-2} \right] \\ &= \left[ n^{-1} \sum_{t=k+1}^n e_{t-1} e_{t-j} - n^{-1} (1-n^{-\delta}\gamma)^{k-j} \sum_{t=k+1}^n e_{t-k} e_{t-1} \right] \\ &\quad + n^{-\delta}\gamma n^{-1} \left[ \sum_{t=k+1}^n X_{t-2} e_{t-j} - (1-n^{-\delta}\gamma)^{k-j} \sum_{t=k+1}^n X_{t-2} e_{t-k} \right]. \end{aligned}$$

Note that

$$\begin{aligned} n^{-1} \sum_{t=k+1}^n X_{t-2} e_{t-k} &= n^{-1} \sum_{t=k+1}^n X_{t-k-1} e_{t-k} + n^{-1} \sum_{t=k+1}^n e_{t-k}^2 + \sum_{\ell=2}^{k-1} n^{-1} \sum_{t=k+1}^n e_{t-\ell} e_{t-k} \\ &= \left[ 1 + n^{-1} \sum_{t=2}^n X_{t-1} e_t \right] + o_p(kn^{-1/2}), \end{aligned}$$

and for  $j \geq 2$ ,

$$\begin{aligned} n^{-1} \sum_{t=k+1}^n X_{t-2} e_{t-j} &= n^{-1} \sum_{t=k+1}^n X_{t-j-1} e_{t-j} + n^{-1} \sum_{t=k+1}^n e_{t-j}^2 + n^{-1} \sum_{\ell=3}^j \sum_{t=k+1}^n e_{t-\ell} e_{t-k} \\ &= \left[ 1 + n^{-1} \sum_{t=2}^n X_{t-1} e_t \right] + o_p(kn^{-1/2}). \end{aligned}$$

Therefore,

$$\begin{aligned} n^{-1} \sum_{t=k+1}^n w_{t-1} Y_{t-1} &= n^{-1} \sum_{t=k+1}^n e_{t-1}^2 + n^{-\delta}\gamma \left[ n^{-1} \sum_{t=2}^n X_{t-1} e_t - (1-n^{-\delta}\gamma)^{k-1} \left( 1 + n^{-1} \sum_{t=2}^n X_{t-1} e_t \right) \right] \\ &\quad + o_p(n^{-1/2}) + o_p(kn^{-0.5-\delta}) \\ &= (1-n^{-\delta}\gamma) + n^{-\delta}\gamma \left[ 1 + n^{-1} \sum_{t=2}^n X_{t-1} e_t \right] \left[ 1 - (1-n^{-\delta}\gamma)^{k-1} \right] \\ &\quad + o_p(n^{-1/2}) + o_p(kn^{-0.5-\delta}), \end{aligned}$$

and for  $j \geq 2$ ,

$$\begin{aligned} n^{-1} \sum_{t=k+1}^n w_{t-j} Y_{t-1} &= n^{-\delta} \gamma \left[ 1 + n^{-1} \sum_{t=2}^n X_{t-1} e_t \right] \left[ 1 - (1 - n^{-\delta} \gamma)^{k-j} \right] \\ &\quad + o_p(n^{-1/2}) + o_p(kn^{-0.5-\delta}) . \end{aligned}$$

Therefore,

$$\begin{aligned} \|n^{-1} \mathbf{b}_k - (1 - n^{-\delta} \gamma) \mathbf{e}_1 - n^{-\delta} \gamma (1 + n^{-1} \sum_{t=1}^n X_{t-1} e_t) (1 - \alpha)\| \\ = o_p(k^{1/2} n^{-1/2}) + o_p(k^{1/2} n^{-0.25-\delta}) . \end{aligned} \quad (\text{A.22})$$

From (A.21) and (A.22), we have

$$\|n^{-1} \mathbf{c}_k\| = o_p(1)$$

and

$$\|n^{-1} \mathbf{b}_k\| = \begin{cases} o_p(k^{1/2} n^{-\delta}) & \text{if } 0 < \delta \leq 1/8 \\ o_p(1) & \text{if } \delta > 1/8 \end{cases} .$$

Therefore, for  $0 < \delta < 0.25$ ,

$$n^{-2+2\delta} \sum_{t=k+1}^n Y_{t-1}^2 - n^{-2+2\delta} \mathbf{b}_k' \mathbf{A}_k^{-1} \mathbf{b}_k = \gamma^2 n^{-2} \sum_{t=2}^n X_{t-1}^2 + o_p(1) . \quad (\text{A.23})$$

Also,

$$\begin{aligned} n^{-1+\delta} \mathbf{b}_k' \mathbf{A}_k^{-1} \mathbf{c}_k &= - \left[ n^\delta (1 - n^{-\delta} \gamma) \mathbf{e}_1' + \gamma (1 + n^{-1} \sum_{t=2}^n X_{t-1} e_t) (1 - \alpha)' \right] (I + \alpha \alpha')^{-1} \mathbf{e}_1 (1 - n^{-\delta} \gamma) + o_p(1) \\ &= -n^\delta (1 - n^{-\delta} \gamma)^2 - \gamma \left[ 1 + n^{-1} \sum_{t=2}^n X_{t-1} e_t \right] + o_p(1) , \end{aligned}$$

and hence

$$\begin{aligned} n^{-1+\delta} \left[ \sum_{t=k+1}^n Y_{t-1} u_t - \mathbf{b}'_k \mathbf{A}_k^{-1} \mathbf{c}_k \right] &= -n^\delta (1-n^{-\delta}\gamma) + n^\delta (1-n^{-\delta}\gamma)^2 + \gamma \left[ 1+n^{-1} \sum_{t=2}^n X_{t-1} e_t \right] + o_p(1) \\ &= \gamma n^{-1} \sum_{t=2}^n X_{t-1} e_t + o_p(1) . \end{aligned} \quad (\text{A.24})$$

Recall that

$$\begin{aligned} s_k^2 &= (n-2k)^{-1} \sum_{t=k+1}^n u_t^2 - n^{-1} \mathbf{c}'_k \mathbf{A}_k^{-1} \mathbf{c}_k - n^{-1} d_k \left[ \sum_{t=k+1}^n Y_{t-1} u_t - \mathbf{b}'_k \mathbf{A}_k^{-1} \mathbf{c}_k \right]^2 \\ &= n^{-1} \sum_{t=1}^n u_t^2 - n^{-1} \boldsymbol{\epsilon}'_1 (\mathbf{I} + \boldsymbol{\alpha} \boldsymbol{\alpha}')^{-1} \boldsymbol{\epsilon}_1 + o_p(1) \\ &\xrightarrow{P} 2-1 = 1. \end{aligned} \quad (\text{A.25})$$

From (A.11), (A.23), (A.24) and (A.25), we get

$$t_k \xrightarrow{D} \Gamma^{-1/2} \xi .$$

(c) Consider the test statistic  $t_{GN}^*$  given in section 2. Let  $\rho^*$  and  $\beta^*$  be initial estimators of  $\rho (=1)$  and  $\beta_n (=n^{-\delta}\gamma-1)$  such that  $\rho^* - 1 = o_p(n^{-1})$ ,  $|\beta^*| < 1$  and  $\beta^* - \beta_n = o_p(n^{-0.5-0.5\delta})$ , for  $0 < \delta < 0.25$ . (It can be shown that the estimator  $\hat{\beta}_r$  considered in Said and Dickey (1985) satisfies this condition.)

Then,

$$n^{-\delta} (1+\beta^*) \xrightarrow{P} \gamma^{-1}$$

and

$$(1+\beta^*)^{-1} (1+\beta_n) \xrightarrow{P} 1 .$$

Following the arguments given in Result A.1, it can be shown that

$$\begin{aligned} (1+\beta^*)V_t^* &= (1+\beta_n)X_{t-1} + f_t \\ &= n^{-\delta}X_{t-1} + f_t \end{aligned}$$

and

$$W_t^* = \sum_{j=0}^{t-2} (-\beta_n)^j e_{t-j} + B_t$$

where

$$f_t = (\beta^* - \beta_n) \left[ e_{t-1} - \sum_{j=1}^{t-2} (-\beta^*)^j e_{t-1-j} \right],$$

$$\begin{aligned} B_t &= (1-\rho^*)\gamma^{-1}n^\delta X_{t-2} + (1-\rho^*)\gamma^{-1}n^\delta \sum_{j=1}^{t-2} e_{t-1-j} (-\beta^*)^{j-1} a_j^{**} \\ &\quad - (\beta^* - \beta_n) \sum_{j=1}^{t-2} j [(-\beta^*)^{j-1} + (-\hat{\beta}^*)^{j-1}] e_{t-1-j}, \end{aligned}$$

$$a_j^{**} = (\beta^* - \beta_n)j + \gamma^{-1}n^\delta \beta^*$$

and  $\hat{\beta}^*$  is a point between  $\beta^*$  and  $\beta_n$ . After some tedious algebra it can be shown that

$$n^{-2} \sum_{t=1}^n V_t^{*2} = n^{-2} \sum_{t=2}^n X_{t-1}^2 + o_p(n^{2\delta-0.5}),$$

$$n^{-1} \sum_{t=1}^n V_t^* e_t = n^{-1} \sum_{t=2}^n X_{t-1} e_t + o_p(n^{2\delta-0.5}),$$

$$n^{-1-\delta} \sum_{t=1}^n W_t^{*2} = (2\gamma)^{-1} + o_p(1),$$

$$n^{-1.5-0.5\delta} \sum_{t=1}^n W_t^* V_t^* = o_p(1),$$

and

$$n^{-0.5-0.5\delta} \sum_{t=1}^n W_t^* e_t = o_p(1).$$

Now, from the arguments used in the proof of Result A.1, it follows that

$$t_{GN} \xrightarrow{D} \Gamma^{-1/2} \xi.$$

(d) To derive the asymptotic distribution of  $Z_\alpha$  in (A.12), consider

$$\begin{aligned} N &= \sum_{t=2}^n Y_{t-1} u_t - \sum_{j=1}^k \sum_{t=j+1}^n u_t u_{t-j} \\ &= \sum_{t=2}^n Y_{t-1} u_t - \sum_{t=2}^k u_t \sum_{j=1}^{t-1} u_{t-j} - \sum_{t=k+1}^n u_t \sum_{j=1}^k u_{t-j} \\ &= \sum_{t=2}^n Y_{t-1} u_t - \sum_{t=2}^k u_t (Y_{t-1} - Y_0) - \sum_{t=k+1}^n u_t (Y_{t-1} - Y_{t-k-1}) \\ &= \sum_{t=2}^n Y_{t-k-1} u_t. \end{aligned}$$

After some routine algebra, it follows that

$$\begin{aligned} N &= \sum_{t=k+1}^n e_{t-k-1} e_t (1 - n^{-\delta} \gamma) - \sum_{t=k+1}^n e_{t-k} e_t \\ &\quad + n^{-2\delta} \gamma^2 \sum_{t=k+1}^n X_{t-k-1} e_t + n^{-\delta} \gamma X_{n-k-1} e_{n-1} + o_p(1). \end{aligned} \quad (\text{A.26})$$

Therefore, for  $0 < \delta < 0.25$ ,

$$n^{-1+2\delta} N = \gamma^2 n^{-1} \sum_{t=2}^n X_{t-1} e_t + o_p(1)$$



and

$$Z_{\alpha} = [n^{-2+2\delta} \sum_{t=2}^n Y_{t-1}^2]^{-1} [n^{-1+2\delta} N]$$

$$\xrightarrow{D} \Gamma^{-1} \xi .$$

(e) To find the distribution of  $Z_{IV}$  given in (A.13), consider

$$\sum_{t=3}^n Y_{t-2} Y_{t-1} = \sum_{t=2}^{n-1} e_{t-1} e_t + \gamma^2 n^{-2\delta} \sum_{t=2}^n X_{t-1}^2 + o_p(n^{-\delta+1}), \quad (A.27)$$

and

$$\sum_{t=3}^n Y_{t-2} u_t = \sum_{t=3}^n e_{t-2} e_t - \sum_{t=2}^{n-1} e_{t-1} e_t + n^{-2\delta} \gamma^2 \sum_{t=2}^n X_{t-1} e_t + o_p(n^{-\delta+0.5}). \quad (A.28)$$

Therefore, from (A.13), (A.27) and (A.28), we get for  $0 < \delta < 0.25$ ,

$$Z_{IV} = [n^{-2+2\delta} \sum_{t=3}^n Y_{t-2} Y_{t-1}]^{-1} [n^{-1+2\delta} \sum_{t=3}^n Y_{t-2} u_t]$$

$$\xrightarrow{D} \Gamma^{-1} \xi . \quad \mathbf{I}$$

The following lemma is used to prove Theorems 3.2 and 3.3.

Lemma A.2: Let  $\{e_t\}$  be a sequence of iid  $(0,1)$  random variables with bounded fourth moments. Consider the random variables,

$$(S_n(r), N_n')' = n^{-1/2} \left( \sum_{t=1}^{[nr]} e_t, \sum_{t=1}^n \epsilon_t^* e_t \right)'$$

where

$$\epsilon_t^* = (e_{t-k-1}, e_{t-k}, e_{t-2}, e_{t-1})',$$

$[nr]$  denotes the integer part of  $nr$ ,  $0 \leq r \leq 1$  and  $k = O(n^{1/4})$ .

Then,

$$(S_n(r), N_n)' \Rightarrow (W(r), N_3, N_2, N_1, N_0)',$$

where  $W(\cdot)$  is a standard Brownian motion,  $N_0, N_1, N_2$  and  $N_3$  are independent  $N(0,1)$  variables independent of  $W(\cdot)$ .

Proof: Let  $\mathcal{F}_t$  denote the  $\sigma$ -field generated by  $\{e_s: s \leq t\}$ . Note that

$(\sum_{t=1}^n \epsilon_t^* e_t, \mathcal{F}_t)$  is a martingale sequence. Also,

$$n^{-1} \sum_{t=k+1}^n \epsilon_t^* \epsilon_t^* \xrightarrow{P} I_4$$

where  $I_4$  denotes an identity matrix of size 4. Since  $e_t$ 's are assumed to have bounded fourth moments, the result follows from Theorem 2.2 of Chan and Wei (1988) and Theorem 3.3 of Helland (1982). **I**

We will now present a proof of Theorem 3.2.

Proof of Theorem 3.2: Recall that the process of  $Y_t$  satisfies

$$Y_t = e_t + n^{-\delta} \gamma X_{t-1}, \quad 0.25 < \delta < 0.5, \quad 0 < \gamma < 2.$$

where  $X_t = \sum_{i=1}^t e_i$ .

(a) From (A.14), (A.15) and (A.16) we get, for  $0.25 < \delta < 0.5$ ,

$$n^{-2+2\delta} \sum_{t=2}^n Y_{t-1}^2 \xrightarrow{P} \gamma^2 \Gamma,$$

$$n^{-1} \sum_{t=2}^n Y_{t-1} u_t \xrightarrow{P} -1$$

and

$$n^{-1} \sum_{t=1}^n u_t^2 \xrightarrow{P} 2 .$$

Therefore, from (A.6) we have

$$n^{-\delta} \hat{\tau} \xrightarrow{P} -[2\gamma^2 \Gamma]^{-1/2} .$$

(b) Note that, from (A.17), (A.21) and (A.22) we have

$$\begin{aligned} n^{-1} \sum_{t=k+1}^n Y_{t-1} u_t - n^{-1} \mathbf{b}_k' \mathbf{A}_k^{-1} \mathbf{c}_k &= -(1-n^{-\delta}\gamma) + (1-n^{-\delta}\gamma)^2 \boldsymbol{\epsilon}_1' (\mathbf{n}^{-1} \mathbf{A}_k)^{-1} \boldsymbol{\epsilon}_1 + o_p(k^{-1}) \\ &= -1 + \boldsymbol{\epsilon}_1' (\mathbf{n}^{-1} \mathbf{A}_k)^{-1} \boldsymbol{\epsilon}_1 + o_p(k^{-1}) . \end{aligned} \quad (\text{A.29})$$

We will now show that

$$k[-1 + \boldsymbol{\epsilon}_1' (\mathbf{n}^{-1} \mathbf{A}_k)^{-1} \boldsymbol{\epsilon}_1] \xrightarrow{P} -1 .$$

Notice that  $\boldsymbol{\epsilon}_1' (\mathbf{n}^{-1} \mathbf{A}_k)^{-1} \boldsymbol{\epsilon}_1$  is the first element of  $(\mathbf{n}^{-1} \mathbf{A}_k)^{-1}$  which is also the inverse of the residual sum of squares in the regression of  $n^{-1/2} w_{t-1}$  on  $n^{-1/2} w_{t-2}, \dots, n^{-1/2} w_{t-k+1}$ , where  $w_{t-i} = e_{t-i} - (1-n^{-\delta}\gamma)^{k-i} e_{t-k}$ . Consider the model

$$n^{-1/2} w_{t-1} = \sum_{i=2}^{k-1} a_i n^{-1/2} w_{t-i} + \text{error} . \quad (\text{A.30})$$

The model (A.30) is equivalent to

$$n^{-1/2} e_{t-1} = \sum_{i=2}^k a_i n^{-1/2} e_{t-i} + \text{error} , \quad (\text{A.31})$$

where

$$\sum_{i=2}^k (1-n^{-\delta}\gamma)^{k-i} a_i = (1-n^{-\delta}\gamma)^{k-1} . \quad (\text{A.32})$$

Therefore, the residual sum of squares of the regression (A.30) is the same as the residual sum of squares of the model (A.31) where the coefficients satisfy the condition (A.32). Let

$$\hat{R}_k = n^{-1} \sum_{t=k+1}^n \hat{v}_{t-1} \hat{v}'_{t-1} ,$$

$$\hat{r}_k = n^{-1} \sum_{t=k+1}^n \hat{v}_{t-1} e_{t-1} ,$$

$$\hat{v}_{t-1} = (e_{t-2}, \dots, e_{t-k})' ,$$

and

$$\hat{a} = \hat{R}_k^{-1} \hat{r}_k .$$

Then, the condition (A.32) is  $\alpha' \hat{a} = (1 - n^{-\delta} \gamma)^k$  where  $\alpha$  is defined in Lemma A.1, and the residual sum of squares of the regression (A.30) is

$$\begin{aligned} [\epsilon_1' (n^{-1} A_k)^{-1} \epsilon_1]^{-1} &= n^{-1} \sum_{t=k+1}^n e_{t-1}^2 - \hat{a}' \hat{R}_k \hat{a} \\ &+ (\alpha' \hat{R}_k^{-1} \alpha)^{-1} [\alpha' \hat{a} - (1 - n^{-\delta} \gamma)^k]^2 . \end{aligned} \quad (A.33)$$

From Berk (1974) we get,

$$\alpha' \hat{a} \xrightarrow{P} 0 ,$$

$$k^{-1} \alpha' \hat{R}_k^{-1} \alpha \xrightarrow{P} 1 ,$$

and

$$\hat{a}' \hat{R}_k \hat{a} = o_p(n^{-1} k) .$$

Now, from (A.33) we have

$$k[\epsilon_1'(n^{-1}A_k)^{-1}\epsilon_1^{-1}] = \frac{k^{-1}(\alpha'\hat{R}_k^{-1}\alpha)k(1-n^{-1}\sum_{t=k+1}^n e_{t-1}^2 + \hat{a}'\hat{R}_k\hat{a}) - [\alpha'\hat{a} - (1-n^{-\delta}\gamma)^{k-1}]^2}{k^{-1}(\alpha'\hat{R}_k^{-1}\alpha)(n^{-1}\sum_{t=k+1}^n e_{t-1}^2 - \hat{a}'\hat{R}_k\hat{a}) - k^{-1}[\alpha'\hat{a} - (1-n^{-\delta}\gamma)^{k-1}]^2}$$

$\xrightarrow{P} -1 .$

(A.34)

From (A.11), (A.23), (A.25), (A.29) and (A.34), we get

$$n^{-\delta}kt_k \xrightarrow{D} -[\gamma^2\Gamma]^{-1/2} .$$

(c) From (A.26) and (A.14) we have, for  $0.25 < \delta < 0.5$ ,

$$n^{-1/2}N = n^{-1/2} \sum_{t=k+1}^n e_{t-k-1}e_t - n^{-1/2} \sum_{t=k+1}^n e_{t-k}e_t + o_p(1)$$

and

$$n^{-2+2\delta} \sum_{t=2}^n Y_{t-1}^2 = \gamma^2 n^{-2} \sum_{t=2}^n X_{t-1}^2 + o_p(1) .$$

Therefore, from Lemma A.2 and (A.12), we have

$$n^{-2\delta+0.5} Z_\alpha \xrightarrow{D} [\gamma^2\Gamma]^{-1} [N_3 - N_2] .$$

(d) From (A.27) and (A.28) we have for  $0.25 < \delta < 0.5$ ,

$$n^{-2+2\delta} \sum_{t=3}^n Y_{t-2}Y_{t-1} = \gamma^2 n^{-2} \sum_{t=3}^n X_{t-1}^2 + o_p(1)$$

and

$$n^{-1/2} \sum_{t=3}^n Y_{t-2}u_t = n^{-1/2} \sum_{t=3}^n e_{t-2}e_t - n^{-1/2} \sum_{t=3}^n e_{t-1}e_t + o_p(1) .$$

Therefore, from Lemma A.2 and (A.13) we get

$$n^{-2\delta+0.5} Z_{IV} \xrightarrow{D} [\gamma^2 \Gamma]^{-1} (N_1 - N_0) . \quad \mathbf{I}$$

Finally, we present a proof for Theorem 3.3

Proof of Theorem 3.3: Here we assume that

$$Y_t = e_t + n^{-\delta} \gamma X_{t-1}, \quad \delta > 0.5, \quad 0 < \gamma < 2$$

where  $X_t = \sum_{i=1}^t e_i$ .

(a) For  $\delta > 0.5$ , note that

$$n^{-1} \sum_{t=2}^n Y_{t-1}^2 \xrightarrow{P} 1 ,$$

$$n^{-1} \sum_{t=2}^n Y_{t-1} u_t \xrightarrow{P} 1 ,$$

$$n^{-1} \sum_{t=2}^n u_t^2 \xrightarrow{P} 2 ,$$

and

$$s_1^2 = n^{-1} \sum_{t=2}^n u_t^2 - [n^{-1} \sum_{t=2}^n Y_{t-1}^2]^{-1} [n^{-1} \sum_{t=2}^n Y_{t-1} u_t]^2$$

$$\xrightarrow{P} 1.$$

Therefore, from (A.6), we get

$$n^{-1/2} \hat{\tau} \xrightarrow{P} -1 .$$

(b) The results (A.29) and (A.34) hold also for  $\delta > 0.5$  and

$$n^{-1} \sum_{t=k+1}^n Y_{t-1} u_t - n^{-1} \mathbf{b}'_k \mathbf{A}_k^{-1} \mathbf{c}_k = -1 + \boldsymbol{\epsilon}'_1 (n^{-1} \mathbf{A}_k)^{-1} \boldsymbol{\epsilon}_1 + o_p(k^{-1})$$

$$= -k^{-1} + o_p(k^{-1}) .$$

Also,

$$\begin{aligned} n^{-1} \mathbf{b}'_k \mathbf{A}_k^{-1} \mathbf{b}_k &= \boldsymbol{\epsilon}'_1 (n^{-1} \mathbf{A}_k)^{-1} \boldsymbol{\epsilon}_1 + o_p(kn^{-1}) \\ &= 1 - k^{-1} + o_p(k^{-1}) . \end{aligned}$$

For  $0.5 < \delta < 0.625$ ,

$$\begin{aligned} n^{-2+2\delta} \sum_{t=k+1}^n Y_{t-1}^2 - n^{-2+2\delta} \mathbf{b}'_k \mathbf{A}_k^{-1} \mathbf{b}_k &= n^{-1+2\delta} n^{-1} \sum_{t=k+1}^n e_{t-1}^2 + \gamma^2 n^{-2} \sum_{t=k+1}^n X_{t-1}^2 \\ &\quad - n^{-1+2\delta} + k^{-1} n^{-1+2\delta} + o_p(1) \end{aligned}$$

$$\xrightarrow{\mathcal{D}} \gamma^2 \Gamma ,$$

$$k[n^{-1} \sum_{t=k+1}^n Y_{t-1} u_t - n^{-1} \mathbf{b}'_k \mathbf{A}_k^{-1} \mathbf{c}_k] \xrightarrow{P} -1 ,$$

and

$$s_k^2 \xrightarrow{P} 1 .$$

Therefore, for  $0.5 < \delta < 0.625$ ,

$$n^{-\delta} k t_k \xrightarrow{\mathcal{D}} -[\gamma^2 \Gamma]^{-1/2} .$$

(c) For  $\delta > 0.625$ ,

$$k[n^{-1} \sum_{t=k+1}^n Y_{t-1}^2 - n^{-1} \mathbf{b}'_k \mathbf{A}_k^{-1} \mathbf{b}_k] = k[1 - 1 + k^{-1} + o_p(k^{-1})]$$

$$\xrightarrow{P} 1 ,$$

$$s_k^2 \xrightarrow{P} 1 ,$$

and

$$k[n^{-1} \sum_{t=k+1}^n Y_{t-1} u_t - n^{-1} b_k' A_k^{-1} c_k] \xrightarrow{P} -1 .$$

Therefore, for  $\delta > 0.625$ ,

$$k^{1/2} n^{-1/2} t_k \xrightarrow{P} -1 .$$

(d) Note that for  $\delta > 0.5$ , from (A.26) we have

$$n^{-1/2} N = n^{-1/2} \sum_{t=k+1}^n e_{t-k-1} e_t - n^{-1/2} \sum_{t=k+1}^n e_{t-k} e_t + o_p(1)$$

and

$$n^{-1} \sum_{t=k+1}^n Y_{t-1}^2 = n^{-1} \sum_{t=k+1}^n e_{t-1}^2 + o_p(1) .$$

Therefore, from (A.12) and Lemma A.2, we get

$$n^{-1/2} Z_{\alpha} \xrightarrow{\mathcal{D}} N_3 - N_2 .$$

(e) For  $0.5 < \delta < 0.75$ , from the arguments used in the proof of Theorem 3.2, we get

$$n^{-2\delta+0.5} Z_{IV} \xrightarrow{\mathcal{D}} [\gamma^2 \Gamma]^{-1} (N_1 - N_0) .$$

(f) For  $\delta > 0.75$ ,

$$\begin{aligned} n^{-1} Z_{IV} + 1 &= \left[ \sum_{t=3}^n Y_{t-2} Y_{t-1} \right]^{-1} \left[ \sum_{t=3}^n Y_{t-2} Y_t \right] \\ &= \left[ n^{-1/2} \sum_{t=3}^n e_{t-2} e_t \right]^{-1} \left[ n^{-1/2} \sum_{t=3}^n e_{t-1} e_t \right] + o_p(1) \\ &\xrightarrow{\mathcal{D}} N_1^{-1} N_0 . \quad \blacksquare \end{aligned}$$



ACKNOWLEDGEMENTS

I am grateful to David Dickey for his comments and suggestions. My thanks also go to the National Science Foundation for research support. Finally, my thanks go to Janice Gaddy for her skill and effort in typing this manuscript.

REFERENCES

- Ahtola, J. A. and G. C. Tiao (1984). Parametric inference for a nearly nonstationary first order autoregressive model. *Biometrika*, **71**, 263-272.
- Berk, K. N. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics*, **2**, 489-502.
- Chan, N. H. and C. Z. Wei (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *The Annals of Statistics*, **15**, 1050-1063.
- Chan, N. H. and C. Z. Wei (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *The Annals of Statistics*, **16**, 367-401.
- Dickey, D. A. and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, **74**, 427-431.
- Fuller, W. A. (1976). *Introduction to Statistical Time Series*, Wiley, New York.
- Hall, A. (1988). Testing for a unit root in the presence of moving average errors. *Biometrika* (forthcoming).
- Helland, I. S. (1982). Central limit theorems for martingales with discrete or continuous time. *Scandinavian Journal of Statistics*, **9**, 79-94.
- Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models. *Biometrika*, **65**, 99-107.

- Phillips, P. C. B. (1987a). Time series regression with a unit root. *Econometrica*, **55**, 277-301.
- Phillips, P. C. B. (1987b). Towards a unified asymptotic theory for autoregression. *Biometrika*, **74**, 535-547.
- Phillips, P. C. B. and P. Perron (1988). Testing for a unit root in time series regression. *Biometrika* (forthcoming).
- Said, S. E. and D. A. Dickey (1984). Testing for unit roots in autoregressive moving average models of unknown order. *Biometrika*, **71**, 599-607.
- Said, S. E. and D. A. Dickey (1985). Hypothesis testing in ARIMA (p,1,q) models. *Journal of the American Statistical Association*, **80**, 369-374.
- Schwert, G. W. (1987). Effects of model specification on tests for unit roots in macroeconomic data. *Journal of Monetary Economics*, **20**, 73-103.
- Schwert, G. W. (1988). Tests for unit roots: A Monte Carlo Investigation. *Journal of Business and Economic Statistics* (forthcoming).