

# On the Asymptotics of ADF Tests for Unit Roots<sup>1</sup>

Yoosoon Chang

*Department of Economics*

*Rice University*

and

Joon Y. Park

*School of Economics*

*Seoul National University*

## Abstract

In this paper, we derive the asymptotic distributions of Augmented-Dickey-Fuller (ADF) tests under very mild conditions. The tests were originally proposed and investigated by Said and Dickey (1984) for testing unit roots in finite-order ARMA models with iid innovations, and are based on a finite AR process of order increasing with the sample size. Our conditions are significantly weaker than theirs. In particular, we allow for general linear processes with martingale difference innovations, possibly having conditional heteroskedasticities. The linear processes driven by ARCH type innovations are thus permitted. The range for the permissible increasing rates for the AR approximation order is also much wider. For the usual  $t$ -type test, we only require that it increase at order  $o(n^{1/2})$  while they assume that it is of order  $o(n^\kappa)$  for some  $\kappa$  satisfying  $0 < \kappa \leq 1/3$ .

*This version:* November, 2001

*Key words and phrases:* ADF tests, unit roots, asymptotics, linear process, autoregressive approximation.

---

<sup>1</sup>We wish to thank the Editor, an Associate Editor and two anonymous referees for useful comments, which have greatly improved the exposition of an earlier version of this paper. Park thanks the Department of Economics at Rice University, where he is an Adjunct Professor, for its continuing hospitality and secretarial support. This research was supported by Korea Research Foundation.

## 1. Introduction

The tests for unit roots in AR processes were first proposed and investigated by Dickey and Fuller (1979, 1981). The tests by Dickey-Fuller (DF) are based on finite-order AR models, the orders of which are assumed to be known.<sup>2</sup> They were later extended by Said and Dickey (1984) to allow for finite ARMA processes of unknown order. They show that the tests, which are frequently referred to as Augmented-Dickey-Fuller (ADF) tests have the same limiting distributions as DF tests for any finite-order ARMA processes with iid innovations, if we increase the order of the approximating AR regression models appropriately as the sample size grows. It seems that ADF tests are most commonly used by practitioners, along with the tests by Phillips (1987) and Phillips and Perron (1988), which will be called PP tests in the paper.

It is widely conjectured that ADF tests are valid for a class of models broader than those considered by Said and Dickey (1984). In particular, it is routinely assumed that they have the same asymptotic distributions as DF tests, when the underlying time series are generated by general linear processes having martingale difference innovations and satisfying mild coefficient summability conditions. In much of the unit root literature, ADF tests are indeed considered under the same conditions as PP tests, whose asymptotics have long been established under much weaker conditions. To the best of our knowledge, however, it has not yet been shown rigorously that ADF tests have the same asymptotically invariant distributions as DF tests for such a wide class of data generating processes.

The range of the increasing rates for the AR approximation order required in Said and Dickey (1984) is also believed to be unnecessarily stringent. They assume that it increases at order  $o(n^\kappa)$ , for some  $\kappa$  satisfying  $0 < \kappa \leq 1/3$ , as the sample size  $n$  gets large. In particular, the logarithmic rate of increase in the order of approximating AR regressions is not permitted. This can be a serious limitation for the practical applications of the tests. In practice, we often use the order selection rules such as AIC and BIC, which set the AR order increasing at a logarithmic rate. Ng and Perron (1995) indeed show that their results hold in the absence of the lower bound, and thereby validate the use of such information criteria.

The purpose of this paper is to verify the validity of ADF tests under a set of sufficient conditions that are flexible enough to include most of the interesting models used in practical econometrics. Our conditions allow for general linear processes driven by martingale difference innovations, which may well have conditional heteroskedasticities as in ARCH processes. We only assume a very mild summability condition on their coefficients. This contrasts with Said and Dickey (1984), who consider linear processes with geometrically decreasing coefficients and iid innovations. Moreover, a much wider range of the increasing rates for the AR approximation order is permitted. For the usual  $t$ -type test, we require that the rate increase at order  $o(n^{1/2})$  in contrast to  $o(n^{1/3})$  assumed in Said and Dickey (1984).

The plan of the paper is as follows. Section 2 introduces the model and assump-

---

<sup>2</sup>Pantula (1986, 1988) extended the tests to AR(1) models with martingale difference errors and to AR( $p$ ) models with ARCH errors.

tions. The ADF tests and their asymptotic results are presented in Section 3. Section 4 concludes the paper, and Section 5 collect all mathematical proofs.

## 2. The Model and Assumptions

In this section, we present the model and assumptions. Let the time series  $(y_t)$  be given by

$$y_t = \alpha y_{t-1} + u_t \quad (1)$$

with  $(u_t)$  generated as

$$u_t = \pi(L)\varepsilon_t \quad (2)$$

where  $\varepsilon_t$  is white noise,  $L$  is the usual lag operator and

$$\pi(z) = \sum_{k=0}^{\infty} \pi_k z^k$$

The test of the unit root null hypothesis  $\alpha = 1$  will be considered for  $(y_t)$  given as in (1), against the alternative of stationarity  $|\alpha| < 1$ . The initial value  $y_0$  of  $(y_t)$  does not affect our subsequent asymptotics as long as  $y_0 = O_p(1)$ , and therefore, we set  $y_0 = 0$  for expositional brevity. We make the following assumptions.

**Assumption 1** Let  $(\varepsilon_t, \mathcal{F}_t)$  be a martingale difference sequence, with some filtration  $(\mathcal{F}_t)$ , such that (a)  $\mathbf{E}(\varepsilon_t^2) = \sigma^2$ , (b)  $(1/n) \sum_{t=1}^n \varepsilon_t^2 \rightarrow_p \sigma^2$  and (c)  $\mathbf{E}|\varepsilon_t|^r < K$  with  $r \geq 4$ , where  $K$  is some constant depending only upon  $r$ .

**Assumption 2** Let  $\pi(z) \neq 0$  for all  $|z| \leq 1$ , and  $\sum_{k=0}^{\infty} |k|^s |\pi_k| < \infty$  for some  $s \geq 1$ .

Our specification in (2) with conditions in Assumptions 1 and 2 allows  $(u_t)$  to be generated as a quite general linear process. Assumption 1 sets the innovation sequence  $(\varepsilon_t)$  to be martingale differences. With the coefficient summability condition in Assumption 2, condition (a) implies that  $(u_t)$  is weakly stationary. Given condition (a), condition (b) requires that the weak law of large numbers hold for the squared innovations  $(\varepsilon_t^2)$ , for which it suffices to assume suitable mixing conditions for them. Under Assumption 2, we have due to the Marcinkiewicz-Zygmund inequality in, e.g., Stout (1974, Theorem 3.3.6), that the  $r$ -th moment of  $(u_t)$  exists and is bounded uniformly in  $t$ . Note that second order stationary ARCH and GARCH models satisfy our conditions in Assumption 1.

The asymptotics developed in the literature for PP and other nonparametric tests allow for unconditional, as well as conditional, heterogeneity. For these tests, we commonly assume that  $(u_t)$  in (2) is generated by martingale difference innovations  $(\varepsilon_t)$  satisfying

$$\frac{1}{n} \sum_{t=1}^n \mathbf{E}(\varepsilon_t^2 | \mathcal{F}_{t-1}) \rightarrow_p \sigma^2 \quad (3)$$

and for some constant  $K$

$$\mathbf{E}(\varepsilon_t^4 | \mathcal{F}_{t-1}) < K \text{ a.s.} \quad (4)$$

which corresponds to our Assumption 1 (in addition to the coefficient summability condition comparable to our Assumption 2) [see, e.g., Stock (1994)]. Conditions in (3) and (4) imply condition (b) in Assumption 1 [see Hall and Heyde (1980, Theorem 2.23)]. Of course, condition (4) alone also implies condition (c) in Assumption 1. Therefore, besides unconditional homogeneity in condition (a), Assumption 1 is weaker than conditions (3) and (4) taken together.

In place of Assumption 1, we also consider

**Assumption 1'** Let  $(\varepsilon_t, \mathcal{F}_t)$  be a martingale difference sequence, with some filtration  $(\mathcal{F}_t)$ , such that (a)  $\mathbf{E}(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma^2$  and (b)  $\mathbf{E}|\varepsilon_t|^r < K$  with  $r \geq 4$ , where  $K$  is some constant depending only upon  $r$ .

Assumption 1' is stronger than Assumption 1. Condition (a) in Assumption 1' implies condition (a) in Assumption 1. Given condition (b) in Assumption 1', condition (a) in Assumption 1' also implies condition (b) in Assumption 1 [see again Hall and Heyde (1980, Theorem 2.23)]. Obviously, condition (3) holds under condition (a) in Assumption 1', and condition (4) implies condition (b) in Assumption 1'. Therefore, conditions (3) and (4) are neither necessary nor sufficient for Assumption 1'.

**Remark 2.1: AR Approximation** Under Assumptions 1 and 2, we may write

$$\alpha(L)u_t = \varepsilon_t$$

with

$$\alpha(z) = 1 - \sum_{k=1}^{\infty} \alpha_k z^k$$

and approximate  $(u_t)$  in  $r$ -th mean by a finite order AR process

$$u_t = \alpha_1 u_{t-1} + \cdots + \alpha_p u_{t-p} + \varepsilon_{p,t}$$

with

$$\varepsilon_{p,t} = \varepsilon_t + \sum_{k=p+1}^{\infty} \alpha_k u_{t-k}$$

It is well known [see, e.g., Brillinger (1975)] that condition in Assumption 2 implies that  $\sum_{k=1}^{\infty} |k|^s |\alpha_k| < \infty$ , and we have  $\sum_{k=p+1}^{\infty} |\alpha_k| = o(p^{-s})$ . Therefore, given the existence of the  $r$ -th moment of  $(u_t)$  implied by Assumptions 1 and 2,

$$\mathbf{E}|\varepsilon_{p,t} - \varepsilon_t|^r \leq \mathbf{E}|u_t|^r \left( \sum_{k=p+1}^{\infty} |\alpha_k| \right)^r = o(p^{-rs})$$

The approximation error thus becomes small as  $p$  gets large. Later, we will rely on the above AR approximation for  $(u_t)$  with  $p$  increasing as the sample size. To make it explicit that  $p$  is a function of sample size  $n$ , we will often write  $p = p_n$ .

**Remark 2.2: Beveridge-Nelson Representation** We may write  $(u_t)$  as

$$u_t = \pi(1)\varepsilon_t + (\bar{u}_{t-1} - \bar{u}_t) \quad (5)$$

where

$$\bar{u}_t = \sum_{k=0}^{\infty} \bar{\pi}_k \varepsilon_{t-k}, \quad \bar{\pi}_k = \sum_{i=k+1}^{\infty} \pi_i$$

Under our condition in Assumption 2, we have  $\sum_{k=0}^{\infty} |\bar{\pi}_k| < \infty$  [see Phillips and Solo (1992)] and therefore  $(\bar{u}_t)$  is well defined both in a.s. and  $L^r$  sense [see Brockwell and Davis (1991)].

Under the unit root hypothesis, we may now sum (5) on both sides getting

$$y_t = \pi(1)w_t + (\bar{u}_0 - \bar{u}_t) \quad (6)$$

where  $w_t = \sum_{k=1}^t \varepsilon_k$ . Consequently,  $(y_t)$  behaves asymptotically as the constant  $\pi(1)$  multiple of  $(w_t)$ . Note that  $(\bar{u}_t)$  is stochastically of smaller order of magnitude than  $(w_t)$ . The representations in (5) and (6) were used originally by Beveridge and Nelson (1981) to decompose aggregate economic time series into permanent and transitory components. They are fully and rigorously developed in Phillips and Solo (1992), and used to obtain asymptotics for linear processes.

**Assumption 3** Let  $p_n \rightarrow \infty$  and  $p_n = o(n^{1/2})$  as  $n \rightarrow \infty$ .

Conditions in Assumptions 2 and 3 are significantly weaker than the ones used by Said and Dickey (1984). They only consider  $(u_t)$  generated by a finite order ARMA process and thus effectively look at the case that  $(\pi_k)$  decays geometrically. Their assumption corresponds to  $s = \infty$  in our Assumption 2. Also, they assume  $p_n = cn^\kappa$  for  $0 < \kappa \leq 1/3$ . Therefore, for instance, the logarithmic rate for  $p$  is not allowed in their result. This may restrict the use of the tests in some practical applications.<sup>3</sup>

In addition to Assumption 3, we also consider

**Assumption 3'** Let  $p_n \rightarrow \infty$  and  $p_n = o((n/\log n)^{1/2})$  as  $n \rightarrow \infty$ .

**Assumption 3''** Let  $p_n \rightarrow \infty$  and  $p_n = o(n^{1/3})$  as  $n \rightarrow \infty$ .

In Assumptions 3' and 3'', we require  $p$  to increase at slower rates.

### 3. ADF Tests and Their Limiting Distributions

The test of the unit root hypothesis for the time series  $(y_t)$  given by (1) and (2) can be based on the regression

$$y_t = \alpha y_{t-1} + \sum_{k=1}^p \alpha_k \Delta y_{t-k} + \varepsilon_{p,t} \quad (7)$$

---

<sup>3</sup>Some statistical packages like SPLUS set by default the maximum lag length to be  $10 \log_{10}(n)$  for the order selection criteria such as AIC and BIC.

due to Remark 2.1 above. Note that under the null hypothesis we have  $\alpha = 1$  and  $\Delta y_t = u_t$ . To introduce the test statistics, we define

$$x_{p,t} = (\Delta y_{t-1}, \dots, \Delta y_{t-p})'$$

and subsequently let

$$A_n = \sum_{t=1}^n y_{t-1} \varepsilon_{p,t} - \left( \sum_{t=1}^n y_{t-1} x'_{p,t} \right) \left( \sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \left( \sum_{t=1}^n x_{p,t} \varepsilon_{p,t} \right) \quad (8)$$

$$B_n = \sum_{t=1}^n y_{t-1}^2 - \left( \sum_{t=1}^n y_{t-1} x'_{p,t} \right) \left( \sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \left( \sum_{t=1}^n x_{p,t} y_{t-1} \right) \quad (9)$$

$$C_n = \sum_{t=1}^n \varepsilon_{p,t}^2 - \left( \sum_{t=1}^n \varepsilon_{p,t} x'_{p,t} \right) \left( \sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \left( \sum_{t=1}^n x_{p,t} \varepsilon_{p,t} \right) \quad (10)$$

Now we have

$$\begin{aligned} \hat{\alpha}_n - 1 &= A_n B_n^{-1} \\ \hat{\sigma}_n^2 &= n^{-1} (C_n - A_n^2 B_n^{-1}) \\ s(\hat{\alpha}_n)^2 &= \hat{\sigma}_n^2 B_n^{-1} \end{aligned}$$

where  $\hat{\alpha}_n$  is the OLS estimator of  $\alpha$ ,  $\hat{\sigma}_n^2$  is the usual error variance estimator, and  $s(\hat{\alpha}_n)$  is the estimated standard error for  $\hat{\alpha}_n$ . We also let

$$\hat{\alpha}_n(1) = 1 - \sum_{k=1}^p \hat{\alpha}_{p,k} \quad (11)$$

where  $\hat{\alpha}_{p,k}$ 's are the OLS estimators of  $\alpha_k$ 's in regression (7).

The statistics that we will consider in the paper are given by

$$T_n = \frac{\hat{\alpha}_n - 1}{s(\hat{\alpha}_n)} \quad (12)$$

$$S_n = \frac{n(\hat{\alpha}_n - 1)}{\hat{\alpha}_n(1)} \quad (13)$$

Note that  $T_n$  is the  $t$ -statistic for the unit root hypothesis, and  $S_n$  is a normalized unit root regression coefficient. The tests based on  $T_n$  and  $S_n$  will be referred to respectively as the  $t$ -test and the coefficient test. They are the ADF tests which are extensions of the tests considered by Dickey and Fuller (1979, 1981) for the AR(1) model. Said and Dickey (1984) looked at  $T_n$  only.

Now we derive the asymptotic null distributions of the statistics  $T_n$  and  $S_n$  defined in (12) and (13). In what follows, we assume that  $\alpha = 1$ . Also, we use the notation  $\|\cdot\|$  to signify the usual Euclidean norm. We define  $\|x\| = (x_1^2 + \dots + x_p^2)^{1/2}$  for a  $p$ -vector  $x = (x_i)$ , and let  $\|A\| = \max_x \|Ax\|/\|x\|$  for a  $p \times p$  matrix  $A$ .

**Lemma 3.1** Under Assumptions 1, 2 and 3, we have for large  $n$

$$\begin{aligned} \text{(a)} \quad & \frac{1}{n} \sum_{t=1}^n y_{t-1} \varepsilon_{p,t} = \pi(1) \frac{1}{n} \sum_{t=1}^n w_{t-1} \varepsilon_t + o_p(1) \\ \text{(b)} \quad & \frac{1}{n^2} \sum_{t=1}^n y_{t-1}^2 = \pi(1)^2 \frac{1}{n^2} \sum_{t=1}^n w_{t-1}^2 + o_p(1) \\ \text{(c)} \quad & \frac{1}{n} \sum_{t=1}^n \varepsilon_{p,t}^2 = \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2 + o_p(p^{-s}) \end{aligned}$$

**Lemma 3.2** Under Assumptions 1, 2 and 3, we have for large  $n$

$$\begin{aligned} \text{(a)} \quad & \left\| \left( \frac{1}{n} \sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \right\| = O_p(1) \\ \text{(b)} \quad & \left\| \sum_{t=1}^n x_{p,t} y_{t-1} \right\| = O_p(np^{1/2}) \\ \text{(c)} \quad & \left\| \sum_{t=1}^n x_{p,t} \varepsilon_{p,t} \right\| = o_p(np^{-1/2}) \end{aligned}$$

The results in Lemma 3.1 are well expected from Remarks 2.1 and 2.2. If we let  $A_n, B_n$  and  $C_n$  be defined as in (8)–(10), then it follows from Lemmas 3.1 and 3.2 that

$$\begin{aligned} n^{-1} A_n &= \pi(1) \frac{1}{n} \sum_{t=1}^n w_{t-1} \varepsilon_t + o_p(1) \\ n^{-2} B_n &= \pi(1)^2 \frac{1}{n^2} \sum_{t=1}^n w_{t-1}^2 + o_p(1) \\ n^{-1} C_n &= \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2 + o_p(1) \end{aligned}$$

since

$$\begin{aligned} & \left| \left( \sum_{t=1}^n y_{t-1} x'_{p,t} \right) \left( \sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \left( \sum_{t=1}^n x_{p,t} \varepsilon_{p,t} \right) \right| \\ & \leq \left\| \sum_{t=1}^n y_{t-1} x'_{p,t} \right\| \left\| \left( \sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \right\| \left\| \sum_{t=1}^n x_{p,t} \varepsilon_{p,t} \right\| = o_p(n) \end{aligned}$$

due to Lemma 3.2.

Under given assumptions, the parameter estimates  $\hat{\sigma}_n^2$  and  $\hat{\alpha}_n(1)$  used to define the statistics  $T_n$  and  $S_n$  are consistent. Moreover, if we let

$$\hat{\beta}_p = (\hat{\alpha}_{p,1}, \dots, \hat{\alpha}_{p,p})', \quad \beta_p = (\alpha_1, \dots, \alpha_p)'$$

then  $\hat{\beta}_p$  is also consistent for  $\beta_p$  under suitable conditions.

**Lemma 3.3** Under Assumptions 1, 2 and 3, we have  $\hat{\sigma}_n^2 \rightarrow_p \sigma^2$  as  $n \rightarrow \infty$ .

**Lemma 3.4** Under Assumptions 1, 2 and 3'', we have  $\|\hat{\beta}_p - \beta_p\| = o_p(p^{-1/2})$  and  $\hat{\alpha}_n(1) = \alpha(1) + o_p(1)$  for large  $n$ .

**Lemma 3.5** Let Assumptions 1', 2 and 3' hold. We have

$$\hat{\beta}_p = \beta_p + O_p((\log n/n)^{1/2}) + o(p^{-s})$$

uniformly for large  $n$ . Moreover, it follows that

$$\hat{\alpha}_n(1) = \alpha(1) + O_p(p(\log n/n)^{1/2}) + o(p^{-s})$$

for large  $n$ .

Consequently, we have under Assumptions 1, 2 and 3

$$T_n = \frac{\frac{1}{n} \sum_{t=1}^n w_{t-1} \varepsilon_t}{\sigma \left( \frac{1}{n^2} \sum_{t=1}^n w_{t-1}^2 \right)^{1/2}} + o_p(1)$$

Moreover, we have under Assumptions 1, 2 and 3'' or under Assumptions 1', 2 and 3'

$$S_n = \frac{\frac{1}{n} \sum_{t=1}^n w_{t-1} \varepsilon_t}{\frac{1}{n^2} \sum_{t=1}^n w_{t-1}^2} + o_p(1)$$

Note that  $\pi(1) = 1/\alpha(1)$  and  $\alpha(1) \neq 0$ . Therefore, if we define  $\hat{\pi}_n(1) = 1/\hat{\alpha}_n(1)$ , then we have  $\hat{\pi}_n(1) \rightarrow_p \pi(1)$  if and only if  $\hat{\alpha}_n(1) \rightarrow_p \alpha(1)$ .

The asymptotic null distributions of  $T_n$  and  $S_n$  can now be easily obtained using the results in Phillips (1987) or Chan and Wei (1988). They are given in the following theorem. Note that, under Assumption 1, the required invariance principle holds for the partial sum process constructed from  $(\varepsilon_t)$ . See Hall and Heyde (1980, Theorem 4.1). We let  $W$  be the standard Brownian motion.

**Theorem 3.6 (Limiting Distributions of ADF Tests)** Under Assumptions 1, 2 and 3, we have

$$T_n \rightarrow_d \frac{\int_0^1 W_t dW_t}{\left( \int_0^1 W_t^2 dt \right)^{1/2}}$$



as  $n \rightarrow \infty$ . Moreover, under Assumptions 1, 2 and 3'' or under Assumptions 1', 2 and 3', we have

$$S_n \rightarrow_d \frac{\int_0^1 W_t dW_t}{\int_0^1 W_t^2 dt}$$

as  $n \rightarrow \infty$ .

The asymptotic null distributions of  $T_n$  and  $S_n$  are thus identical to those of the corresponding statistics studied in Dickey and Fuller (1979, 1981). They are tabulated in Fuller (1995).

Our result for  $T_n$  only requires that  $p = o(n^{1/2})$ , contrastingly with the more stringent condition  $p = o(n^{1/3})$  used in Berk (1974). This relaxation is possible, since  $T_n$  does not involve any estimates of the coefficients ( $\alpha_k$ ) of the lagged difference terms. Our condition  $p = o(n^{1/2})$  in Assumption 3 is not sufficient for the consistency of these estimates. This can be a serious limitation, since the lag length is often selected using some hypothesis tests on these coefficients. To validate such procedures, we must have slower increasing rates for  $p$  as we specify in Assumptions 3' or 3''. The rate  $p = o((n/\log n)^{1/2})$  is sufficient for the models with homogeneous martingale difference innovations, while we must have  $p = o(n^{1/3})$  as in Berk (1974) for more general models with possibly heterogenous martingale difference innovations. Indeed, these are the conditions that we impose to get the asymptotics of  $S_n$ , which includes the estimate of  $\alpha(1)$ .

**Remark 3.1: Models with Deterministic Trends** The models with deterministic trends can be analyzed similarly. If the time series  $(z_t)$  is given by

$$z_t = \mu + y_t \quad \text{or} \quad z_t = \mu + \tau t + y_t \tag{14}$$

and  $(y_t)$  is generated as in (1), the unit root hypothesis can be tested in regression (7) using residuals obtained from the preliminary regression (14). Their distributions are given similarly as those in Theorem 3.4, respectively with demeaned and detrended Brownian motions

$$W_t^\mu = W_t - \int_0^1 W_s ds, \quad W_t^\tau = W_t + (6t-4) \int_0^1 W_s ds - (12t-6) \int_0^1 s W_s ds$$

in place of standard Brownian motion  $W$ . Though we do not report the details, our results here can easily be extended to obtain the asymptotic theory for the unit root tests in models involving deterministic trends.

**Remark 3.2: Near Unit Root Models** Asymptotics for the near unit root models can also easily be obtained. If we use local-to-unity formulation for  $\alpha$  and let

$$\alpha = 1 - \frac{c}{n}$$

with some constant  $c > 0$ , then the limiting distributions of ADF tests are given as the same as those in Theorem 3.4 with the standard Brownian motion  $W$  replaced by Ornstein-Uhlenbeck process  $W_c$ , which is given by

$$W_c(t) = \int_0^t \exp[-c(t-s)] dW(s)$$

This can also be shown similarly given our assumptions.

## 4. Conclusion

In this paper, we provide rigorous derivations of the asymptotics for ADF tests of unit roots. The required conditions are flexible enough to include a wide class of unit root models generated by very general time series models: linear processes driven by martingale differences with coefficients decaying at polynomial orders. Moreover, our results are obtained under a minimal assumption on the increasing order for the approximating autoregressions.

## 5. Mathematical Proofs

**Proof of Lemma 3.1** For Part (a), write

$$\sum_{t=1}^n y_{t-1} \varepsilon_{p,t} = \sum_{t=1}^n y_{t-1} \varepsilon_t + \sum_{t=1}^n y_{t-1} (\varepsilon_{p,t} - \varepsilon_t)$$

We have

$$\begin{aligned} \sum_{t=1}^n y_{t-1} \varepsilon_t &= \pi(1) \sum_{t=1}^n w_{t-1} \varepsilon_t + \bar{u}_0 \sum_{t=1}^n \varepsilon_t - \sum_{t=1}^n \bar{u}_{t-1} \varepsilon_t \\ &= \pi(1) \sum_{t=1}^n w_{t-1} \varepsilon_t + O_p(n^{1/2}) \end{aligned}$$

It therefore suffices to show that

$$\begin{aligned} &\sum_{t=1}^n y_{t-1} (\varepsilon_{p,t} - \varepsilon_t) \\ &= \pi(1) \sum_{t=1}^n w_{t-1} (\varepsilon_{p,t} - \varepsilon_t) + \bar{u}_0 \sum_{t=1}^n (\varepsilon_{p,t} - \varepsilon_t) - \sum_{t=1}^n \bar{u}_{t-1} (\varepsilon_{p,t} - \varepsilon_t) \\ &= R_{1n} + R_{2n} + R_{3n} = o_p(n) \end{aligned} \tag{15}$$

to deduce the stated result.

To show (15), we first write

$$\varepsilon_{p,t} - \varepsilon_t = \sum_{k=p+1}^{\infty} \alpha_k u_{t-k} = \sum_{k=p+1}^{\infty} \pi_{p,k} \varepsilon_{t-k}$$

where

$$\sum_{k=p+1}^{\infty} \pi_{p,k}^2 \leq c \sum_{k=p+1}^{\infty} \alpha_k^2 = o(p^{-2s})$$

as in Berk (1974, Proof of Lemma 2, p492). Also, denote by  $\delta_{ij}$  the usual Kronecker delta. To show that  $R_{1n} = o_p(n)$ , we write

$$\begin{aligned} \sum_{t=1}^n w_{t-1}(\varepsilon_{p,t} - \varepsilon_t) &= \sum_{t=1}^n \sum_{i=1}^{t-1} \varepsilon_i \sum_{j=p+1}^{\infty} \pi_{p,j} \varepsilon_{t-j} \\ &= \sum_{j=p+1}^{\infty} \pi_{p,j} \sum_{t=1}^n \sum_{i=1}^{t-1} \varepsilon_{t-i} \varepsilon_{t-j} \\ &= \sigma^2 \sum_{k=p+1}^{n-1} (n-k) \pi_{p,k} + \sum_{j=p+1}^{\infty} \pi_{p,j} \sum_{t=1}^n \sum_{i=1}^{t-1} (\varepsilon_{t-i} \varepsilon_{t-j} - \sigma^2 \delta_{ij}) \end{aligned}$$

We have

$$\left| \sum_{k=p+1}^{n-1} (n-k) \pi_{p,k} \right| \leq n \sum_{k=p+1}^{\infty} |\pi_{p,k}| = o(np^{-s})$$

Moreover, we have

$$\begin{aligned} &\left( \mathbf{E} \left[ \sum_{j=p+1}^{\infty} \pi_{p,j} \sum_{t=1}^n \sum_{i=1}^{t-1} (\varepsilon_{t-i} \varepsilon_{t-j} - \sigma^2 \delta_{ij}) \right]^2 \right)^{1/2} \\ &\leq \sum_{j=p+1}^{\infty} |\pi_{p,j}| \left( \mathbf{E} \left[ \sum_{t=1}^n \sum_{i=1}^{t-1} (\varepsilon_{t-i} \varepsilon_{t-j} - \sigma^2 \delta_{ij}) \right]^2 \right)^{1/2} \end{aligned}$$

which is bounded by

$$c n K^{1/2} \left( \sum_{j=p+1}^{\infty} |\pi_{p,j}| \right) = o(np^{-s})$$

for some constant  $c$ . It now follows immediately that  $R_{1n} = o_p(n)$ .

To deduce that  $R_{2n} = o_p(n)$ , we simply note that

$$\sum_{t=1}^n (\varepsilon_{p,t} - \varepsilon_t) = \sum_{k=p+1}^{\infty} \pi_{p,k} \sum_{t=1}^n \varepsilon_{t-k} = o_p(n^{1/2} p^{-s})$$

which follows similarly as above. Finally, to show that  $R_{3n} = o_p(n)$ , we write

$$\begin{aligned} \sum_{t=1}^n \bar{u}_{t-1}(\varepsilon_{p,t} - \varepsilon_t) &= \sum_{i=0}^{\infty} \sum_{j=p+1}^{\infty} \bar{\pi}_i \pi_{p,j} \sum_{t=1}^n \varepsilon_{t-i-1} \varepsilon_{t-j} \\ &= n \sigma^2 \sum_{k=p+1}^{\infty} \bar{\pi}_{k-1} \pi_{p,k} \\ &\quad + \sum_{i=0}^{\infty} \sum_{j=p+1}^{\infty} \bar{\pi}_i \pi_{p,j} \sum_{t=1}^n (\varepsilon_{t-i-1} \varepsilon_{t-j} - \sigma^2 \delta_{i+1,j}) \end{aligned}$$

One may easily see that the first term is of order  $o(np^{-s})$  and that the second term is of order  $o_p(n^{1/2}p^{-s})$ , using the same arguments as above.

The result in Part (b) follows immediately from Phillips and Solo (1992). Moreover, the result stated in Part (c) can easily be deduced from the inequality

$$\left| \left( \frac{1}{n} \sum_{t=1}^n \varepsilon_{p,t}^2 \right)^{1/2} - \left( \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2 \right)^{1/2} \right| \leq \left[ \frac{1}{n} \sum_{t=1}^n (\varepsilon_{p,t} - \varepsilon_t)^2 \right]^{1/2}$$

and the fact that

$$\begin{aligned} \mathbf{E} \left[ \frac{1}{n} \sum_{t=1}^n (\varepsilon_{p,t} - \varepsilon_t)^2 \right] &= \mathbf{E} \left[ \frac{1}{n} \sum_{t=1}^n \left( \sum_{k=p+1}^{\infty} \pi_{p,k} \varepsilon_{t-k} \right)^2 \right] \\ &= \sigma^2 \sum_{k=p+1}^{\infty} \pi_{p,k}^2 = o(p^{-2s}) \end{aligned}$$

Note that  $(1/n) \sum_{t=1}^n \varepsilon_t^2 = O_p(1)$ .

**Proof of Lemma 3.2** We use various results in Berk (1974) in the proof. Though his results were derived under the iid assumption for  $(\varepsilon_t)$ , the results cited here hold under our conditions (a) and (c) in Assumption 1 and Assumption 2, which imply second-order stationarity and uniform boundedness of the fourth moments for  $(u_t)$  and  $(\varepsilon_t)$ .

To show the result in Part (a), we let  $\Gamma_k = \mathbf{E}(u_t u_{t-k})$  be the autocovariance function of  $(u_t)$ , and define

$$_{pp} = (\Gamma_{i-j})_{i,j=1}^p$$

Then it follows from Berk (1974, Proof of Lemma 3, p493) that

$$\mathbf{E} \left\| \left( \frac{1}{n} \sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} - \frac{-1}{_{pp}} \right\|^2 \leq c n^{-1} p^2$$

for some constant  $c$ . Therefore,

$$\left\| \left( \frac{1}{n} \sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} - \frac{-1}{_{pp}} \right\| = O_p(n^{-1/2} p)$$

Moreover, as is well known [see, e.g., Berk (1974, Equation(2.14), p493)],

$$\left\| \frac{-1}{_{pp}} \right\| = O(1)$$

for all  $p$ . The result stated in Part (a) now follows readily, since

$$\left\| \left\| \left( \frac{1}{n} \sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \right\| - \left\| \frac{-1}{_{pp}} \right\| \right\| \leq \left\| \left( \frac{1}{n} \sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} - \frac{-1}{_{pp}} \right\|$$

The proof for Part (a) is therefore complete.

To show the result in Part (b), we let  $(\Gamma_k)$  be the autocovariance function of  $(u_t)$  and use the fact

$$\mathbf{E} \left[ \sum_{t=1}^n (u_{t-i} u_{t-j} - \Gamma_{i-j}) \right]^2 = O(n)$$

which holds uniformly in  $i$  and  $j$ . See, e.g., Berk (1974, Equations (2.10) and (2.11), page 491). In what follows, we let  $y_t = 0$  for all  $t \leq 0$  by convention.

Let  $1 \leq j \leq p$  and write

$$\sum_{t=1}^n y_{t-1} u_{t-j} = \sum_{t=1}^n y_{t-1} u_t + R_n$$

where

$$R_n = \sum_{t=1}^n y_{t-1} u_{t-j} - \sum_{t=1}^n y_{t-1} u_t$$

We will show that  $R_n = O_p(n)$  uniformly in  $j$ ,  $1 \leq j \leq p$ . First note that we have for each  $j = 1, \dots, p$

$$\sum_{t=1}^n y_{t-1} u_t = \sum_{t=1}^n y_{t-j-1} u_{t-j} + \sum_{t=n-j+1}^n y_{t-1} u_t$$

and we may rewrite  $R_n$  as

$$R_n = \sum_{t=1}^n (y_{t-1} - y_{t-j-1}) u_{t-j} - \sum_{t=n-j+1}^n y_{t-1} u_t = R_{1n} - R_{2n}$$

say.

We have

$$\begin{aligned} R_{1n} &= \sum_{t=1}^n \left( \sum_{i=1}^j u_{t-i} \right) u_{t-j} \\ &= n \left( \sum_{i=1}^j \Gamma_{i-j} \right) + \sum_{i=1}^j \left[ \sum_{t=1}^n (u_{t-i} u_{t-j} - \Gamma_{i-j}) \right] \\ &= O(n) + O_p(n^{1/2} p) \end{aligned}$$

uniformly in  $j$ ,  $1 \leq j \leq p$ . Moreover, if we write

$$\begin{aligned} R_{2n} &= \sum_{t=n-j+1}^n \left( \sum_{i=1}^{t-1} u_{t-i} \right) u_t \\ &= \sum_{t=n-j+1}^n \sum_{i=1}^{n-j} u_t u_{t-i} + \sum_{i=n-j+2}^n \sum_{t=n-j+1}^{t-1} u_t u_{t-i} \\ &= R_{2n}^a + R_{2n}^b \end{aligned}$$

then it follows that

$$R_{2n}^a = j \left( \sum_{i=1}^{n-j} \Gamma_i \right) + \sum_{t=n-j+1}^n \left[ \sum_{i=1}^{n-j} (u_t u_{t-i} - \Gamma_i) \right] = O(p) + O_p(n^{1/2}p)$$

and that

$$R_{2n}^b = (j-1) \sum_{i=n-j+1}^{t-1} \Gamma_i + \sum_{t=n-j+2}^n \left[ \sum_{i=n-j+1}^{t-1} (u_t u_{t-i} - \Gamma_i) \right] = O(p) + O_p(p^{3/2})$$

uniformly in  $j$ ,  $1 \leq j \leq p$ . The stated result follows immediately, since  $\sum_{t=1}^n y_{t-1} u_t = O_p(n)$ .

Part (c) readily follows from Berk (1974). Note that

$$\left\| \sum_{t=1}^n x_{p,t} \varepsilon_{p,t} \right\| \leq \left\| \sum_{t=1}^n x_{p,t} (\varepsilon_{p,t} - \varepsilon_t) \right\| + \left\| \sum_{t=1}^n x_{p,t} \varepsilon_t \right\|$$

It follows from Berk (1974, Equation (2.13), p492) that

$$\mathbf{E} \left\| \sum_{t=1}^n x_{p,t} (\varepsilon_{p,t} - \varepsilon_t) \right\|^2 \leq c n^2 p \left( \sum_{k=p+1}^{\infty} \alpha_k^2 \right) = o(n^2 p^{1-2s})$$

Moreover, due to Berk (1974, above Equation (2.17), p493), we have

$$\mathbf{E} \left\| \sum_{t=1}^n x_{p,t} \varepsilon_t \right\|^2 = n p \sigma^2 \mathbf{E} u_t^2 = O(np)$$

To complete the proof, note that

$$n^2 p^{1-2s}, np = o(n^2 p^{-1})$$

since  $p \rightarrow \infty$ ,  $p = o(n^{1/2})$  and  $s \geq 1$  under our assumptions.

**Proof of Lemma 3.3** We have

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2 + o_p(1)$$

from Lemmas 3.1, 3.2 and the subsequent discussions, and the consistency of  $\hat{\sigma}_n^2$  follows directly from (b) in Assumption 1. This completes the proof.

**Proof of Lemma 3.4** We write

$$\hat{\beta}_p - \beta_p = \left( \sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \sum_{t=1}^n x_{p,t} \varepsilon_{p,t}$$

and the stated result for  $\hat{\beta}_p$  follows immediately from

$$\|\hat{\beta}_p - \beta_p\| \leq \left\| \left( \sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \right\| \left\| \sum_{t=1}^n x_{p,t} \varepsilon_{p,t} \right\|$$

due to parts (a) and (c) of Lemma 3.2. To prove the result for  $\hat{\alpha}_n(1)$ , we let  $\iota_p$  be the  $p$ -dimensional vector of ones. It follows that

$$\hat{\alpha}_n(1) = 1 - \iota'_p \hat{\beta}_p$$

Furthermore,

$$\alpha(1) = 1 - \left( \iota'_p \beta_p + \sum_{k=p+1}^{\infty} \alpha_k \right) = \left( 1 - \iota'_p \beta_p \right) + o(p^{-s})$$

as we mentioned in Remark 2.1. Therefore,

$$|\hat{\alpha}_n(1) - \alpha(1)| \leq |\iota'_p \hat{\beta}_p - \iota'_p \beta_p| + o(p^{-s}) \leq p^{1/2} \|\hat{\beta}_p - \beta_p\| + o(p^{-s})$$

from which the stated result for  $\hat{\alpha}_n(1)$  follows immediately.

**Proof of Lemma 3.5** We first write  $(u_t)$  as

$$u_t = \alpha_{p,1} u_{t-1} + \cdots + \alpha_{p,p} u_{t-p} + e_{p,t}$$

where the coefficients  $(\alpha_{p,k})$  are defined so that  $(e_{p,t})$  are uncorrelated with  $(u_{t-k})$  for  $k = 1, \dots, p$ . Moreover, we define

$$u_t = \tilde{\alpha}_{p,1} u_{t-1} + \cdots + \tilde{\alpha}_{p,p} u_{t-p} + \tilde{\varepsilon}_{p,t}$$

It follows from Hamman and Kavalieris (1986, Theorem 2.1) and Bühlmann (1995, Proof of Theorem 3.1) that

$$\max_{1 \leq k \leq p} |\tilde{\alpha}_{p,k} - \alpha_{p,k}| = O((\log n/n)^{1/2}) \text{ a.s.} \quad (16)$$

$$\sum_{k=1}^p |\alpha_{p,k} - \alpha_k| \leq c \sum_{k=p+1}^{\infty} |\alpha_k| = o(p^{-s}) \quad (17)$$

where  $c$  is some constant. Since

$$|\tilde{\alpha}_{p,k} - \alpha_k| \leq |\tilde{\alpha}_{p,k} - \alpha_{p,k}| + |\alpha_{p,k} - \alpha_k|$$

it follows that

$$\max_{1 \leq k \leq p} |\tilde{\alpha}_{p,k} - \alpha_k| = O((\log n/n)^{1/2}) + o(p^{-s}) \text{ a.s.}$$

and that

$$\begin{aligned} |\tilde{\alpha}_n(1) - \alpha(1)| &\leq \sum_{k=1}^p |\tilde{\alpha}_{p,k} - \alpha_{p,k}| + \sum_{k=1}^p |\alpha_{p,k} - \alpha_k| + \sum_{k=p+1}^{\infty} |\alpha_k| \\ &= O(p(\log n/n)^{1/2}) + o(p^{-s}) \text{ a.s.} \end{aligned}$$

from (16) and (17).

However, from the application of simple least squares algebra, we have

$$\begin{pmatrix} \hat{\alpha}_{p,1} \\ \vdots \\ \hat{\alpha}_{p,p} \end{pmatrix} = \begin{pmatrix} \tilde{\alpha}_{p,1} \\ \vdots \\ \tilde{\alpha}_{p,p} \end{pmatrix} - (\hat{\alpha}_n - 1) \left( \sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \left( \sum_{t=1}^n x_{p,t} y_{t-1} \right)$$

and we may easily deduce

$$\hat{\alpha}_n(1) = \tilde{\alpha}_n(1) + O_p(n^{-1}p)$$

Note that we have from parts (a) and (b) of Lemma 3.2 that

$$\begin{aligned} &\left| \iota'_p \left( \sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \left( \sum_{t=1}^n x_{p,t} y_{t-1} \right) \right| \\ &\leq \|\iota_p\| \left\| \left( \sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \right\| \left\| \sum_{t=1}^n x_{p,t} y_{t-1} \right\| = O_p(p) \end{aligned}$$

where  $\iota_p$  is the  $p$ -vector of ones, and  $\hat{\alpha}_n - 1 = O_p(n^{-1})$ . The stated results now follow immediately.

**Proof of Theorem 3.6** Obvious and omitted.

## References

- Berk, K.N. (1974). “Consistent autoregressive spectral estimates,” *Annals of Statistics* 2, 489-502.
- Beveridge, S. and C.R. Nelson (1981). “A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the ‘business cycle’,” *Journal of Monetary Economics* 7, 151-174.
- Brillinger, D.R. (1975). *Time Series: Data Analysis and Theory*. Holt, Rinehart and Winston: New York.
- Bühlmann, P. (1995). “Moving-average representation of autoregressive approximations,” *Stochastic Processes and their Applications*, 60, 331-342.



- Chan, N.H. and C.Z. Wei (1988). "Limiting distributions of least squares estimates of unstable autoregressive processes," *Annals of Statistics* 16, 367-401.
- Dickey, D. A. and W. A. Fuller (1979). "Distribution of estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association* 74, 427-431.
- Dickey, D.A. and W.A. Fuller (1981). "Likelihood ratio statistics for autoregressive time series with a unit root," *Econometrica* 49, 1057-1072.
- Fuller, W.A. (1995). *Introduction to Statistical Time Series*, 2nd ed. Wiley: New York.
- Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and Its Application*. Academic Press: New York.
- Hannan, E.J. and L. Kavalieris (1986). "Regression, autoregression models," *Journal of Time Series Analysis*, 7, 27-49.
- Ng, S. and P. Perron (1995). "Unit root tests in ARMA models with data dependent methods for selection of the truncation lag," *Journal of the American Statistical Association* 90, 268-281.
- Pantula, S.G. (1986). "On asymptotic properties of the least squares estimators for autoregressive time series with a unit root," *Sankhyā*, Series A, 48, 208-218.
- Pantula, S.G. (1988). "Estimation of autoregressive models with ARCH errors," *Sankhyā*, Series B, 50, 119-138.
- Phillips, P.C.B. (1987). "Time series regression with a unit root," *Econometrica* 55, 277-301.
- Phillips, P.C.B. and P. Perron (1988). "Testing for a unit root in time series regression," *Biometrika* 75, 335-346.
- Phillips, P.C.B. and V. Solo (1992). "Asymptotics for linear processes," *Annals of Statistics* 20, 971-1001.
- Said, S.E. and D.A. Dickey (1984). "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika* 71, 599-608.
- Stock, J. H. (1994). "Unit roots, structural breaks and trends," in R.F. Engle and D.L. McFadden (eds), *Handbook of Econometrics*, Vol. IV, pp. 2739-2841, Elsevier, Amsterdam.
- Stout, W.F. (1974). *Almost Sure Convergence*. Academic Press: New York.