CrossMark

REGULAR ARTICLE

Lag truncation and the local asymptotic distribution of the ADF test for a unit root

Emre Aylar[1] · **Stephan Smeekes[2]** ·
Joakim Westerlund[3,4]

Abstract The issue of lag selection in ADF unit root testing is important, even asymptotically, for if the number of lags is not allowed to increase at a certain rate the test might not be correctly sized. However, size control is not the only concern. Indeed, simulations have repeatedly shown how increasing lag lengths tend to be associated with reductions in power, thus adding to the well-known low power problem when the alternative is local to the unit root. But while the simulation evidence is plentiful, there is as of yet almost no asymptotic results that can be used to ascertain whether lag length has any effect on the local asymptotic power of the ADF test. The purpose of the present paper is to fill this gap in the literature.

1 Introduction

The augmented Dickey–Fuller (ADF) unit root test is the most popular of its kind, with countless applications. An issue that arises with the application of this test is the selection of the order of the lag augmentation, p. There are two considerations. On the

✉ Joakim Westerlund
  joakim.westerlund@nek.lu.se

[1] Lund University, Lund, Sweden

[2] Maastricht University, Maastricht, The Netherlands

[3] Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden

[4] Centre for Financial Econometrics, Deakin University, Burwood, Australia

one hand, for the test be correctly sized in the presence of general ARMA errors it is important that p is allowed to increase with the size of the sample, T (see, for example, Said and Dickey 1984). The rate of increase is also important, for only if the rate is fast enough can one rely on conventional data-driven lag selection procedures, such as information criteria (see Ng and Perron 1995; Chang and Park 2002). On the other hand, Monte Carlo evidence indicates that larger values of p are generally associated with reduced power (see Lopez 1997; Ng and Perron 1995; Ng and Perron 2001). Interestingly, while low power is one of the most well-known problems of the ADF test, as far as we are aware no one has as of yet derived any asymptotic power results for the case when p is allowed to increase with T. In fact, most studies, such as those of Said and Dickey (1984), Chang and Park (2002), and Xiao and Phillips (1998), only report the asymptotic distribution under the unit root null hypothesis, although there is typically some conjecture about the behaviour under the alternative that the largest AR root is local-to-unity (see Chang and Park 2002; Xiao and Phillips 1998).[1] The only exceptions known to us are Ng and Perron (2001), whose results are designed specifically to the case when the errors follow a first-order MA process with a root that is local to -1, and Paparoditis and Politis (2017), where the alternative is taken to be that the process is stationary. Both studies confirm that p is important, even asymptotically, and that it can in fact dominate the asymptotic behaviour of the ADF test.

In the present paper, we take the discussion of the last paragraph as our starting point. The purpose is to evaluate the local asymptotic distribution of the ADF test when the errors follow a general linear process driven by martingale difference innovations, which may exhibit conditional heteroskedasticity. The study may therefore be thought of as a local power extension of the study of Chang and Park (2002), who derived the asymptotic null distribution of the ADF test under the same assumption on the errors.

Notation: L is the lag operator, $\to_p$, $\to_w$ and $=_d$ signify convergence in probability, weak convergence, and equality in distribution, respectively, and $\|A\| = \sqrt{\operatorname{tr}(A'A)}$ is the Frobenius norm of any matrix A.

2 Model

The data generating process (DGP) of y_t is the same as in Chang and Park (2002), and is given by

$$y_t = \alpha y_{t-1} + u_t, \tag{1}$$
$$u_t = \pi(L)\varepsilon_t, \tag{2}$$

where $y_0 = 0$, and ε_t and $\pi(L) = \sum_{k=0}^{\infty} \pi_k L^k$ satisfy Assumptions 1 and 2, respectively.

Assumption 1 $(\varepsilon_t, \mathcal{F}_t)$ *is a martingale difference sequence with some filtration* $(\mathcal{F}_t)$, $\mathbf{E}(\varepsilon_t^2) = \sigma^2$, $T^{-1}\sum_{t=1}^{T}\varepsilon_t^2 \to_p \sigma^2$ *and* $\mathbf{E}(|\varepsilon_t|^4) < \infty$.

---

[1] Stock (1991) considers a finite order AR model, the order of which is assumed to be known, and derives the local asymptotic distribution of the ADF test. However, this result holds only for the specific model considered with the restrictive assumption of a known autoregressive order.

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

Remark 1 Assumptions 1 and 2 are the same as in Chang and Park (2002), and are not very restrictive. The assumption that $y_0 = 0$ is more restrictive than necessary, and can be relaxed, provided that $y_0 = O_p(1)$. The fact that there are no deterministic constant and trend terms is restrictive, but as we discuss later in Remark 3 the analysis can be easily extended to accommodate such terms. Note also that the initialization becomes irrelevant if the DGP contains (at least) a constant.

All the results of Chang and Park (2002) are derived under the unit root restriction that $\alpha = 1$. The main contribution of the present paper is to investigate the effect of a violation of this restriction. The particular assumption that we are going to be working under is given by Assumption 3.

Assumption 3 $\alpha = 1 + cT^{-1}$, *where* $c \leq 0$.

As in Chang and Park (2002), $\pi(L)$ has the Beveridge–Nelson (BN) decomposition $\pi(L) = \pi(1) - (1 - L)\bar{\pi}(L)$, where $\bar{\pi}(L) = \sum_{k=0}^{\infty} \bar{\pi}_k L^k$ and $\bar{\pi}_k = \sum_{i=k+1}^{\infty} \pi_i$ (see Phillips and Solo 1992, Lemma 2.3). We can therefore write

$$u_t = \pi(1)\varepsilon_t - \Delta\bar{u}_t, \tag{3}$$

where $\bar{u}_t = \sum_{k=0}^{\infty} \bar{\pi}_k \varepsilon_{t-k}$. Assumption 3 implies

$$y_t = \sum_{k=1}^{t} \alpha^{t-k} u_k = \pi(1)w_t - r_t, \tag{4}$$

where $w_t = \sum_{k=1}^{t} \alpha^{t-k}\varepsilon_k$ and $r_t = \sum_{k=1}^{t} \alpha^{t-k}\Delta\bar{u}_k$.

Under Assumptions 1 and 2, $\pi(L)$ can be inverted, giving

$$\theta(L)u_t = \varepsilon_t, \tag{5}$$

where $\theta(L) = \pi(L)^{-1} = 1 - \sum_{k=1}^{\infty} \theta_k L^k$ (see Chang and Park 2002). The purpose of this paper is to investigate the effect when this infinite-order AR process is truncated at lag p. Let us therefore define $\delta_p(L) = \sum_{k=1}^{p} \theta_k L^{k-1}$, $\delta^p(L) = \sum_{k=p+1}^{\infty} \theta_k L^{k-1}$ and $\delta(L) = \delta_p(L) + \delta^p(L)$, such that $\theta(L) = 1 - \delta(L)L$. In this notation,

$$u_t = \delta_p(L)u_{t-1} + \varepsilon_{p,t}, \tag{6}$$

where

$$\varepsilon_{p,t} = \varepsilon_t + \delta^p(L)u_{t-1}. \tag{7}$$

By using this and the fact that $u_t = y_t - \alpha y_{t-1} = \Delta y_t - (\alpha - 1)y_{t-1}$, we obtain the following equation for y_t:

$$y_t = \alpha y_{t-1} + \delta_p(L)\Delta y_{t-1} - \delta_p(L)(\alpha - 1)y_{t-2} + \varepsilon_{p,t}. \tag{8}$$

At this point, it would seem natural given the approach of Chang and Park (2002) to take $\alpha y_{t-1} + \delta_p(L)\Delta y_{t-1}$ as the approximating regression function, and $\varepsilon_{p,t} - \delta_p(L)(\alpha - 1)y_{t-2}$ as the approximation error. But while this is indeed a possibility, there is a much more elegant approach. To fix ideas, let us write the regression model to be estimated by ordinary least squares (OLS) as

$$y_t = \beta y_{t-1} + \beta_p(L)\Delta y_{t-1} + e_{p,t}, \tag{9}$$

where β and $\beta_p(L)$ are reduced form coefficients, and $e_{p,t}$ is a reduced form error term. We now write these reduced form quantities in terms of the components of the DGP. We begin by noting that

$$- \delta_p(L)(\alpha - 1)y_{t-2} = \delta_p(L)(\alpha - 1)\Delta y_{t-1} - \delta_p(L)(\alpha - 1)y_{t-1}. \tag{10}$$

Consider the last term on the right. Similarly to the BN decomposition for infinite polynomials, we may decompose $\delta_p(L) = \delta_p(1) - (1 - L)\bar{\delta}_p(L)$, where $\bar{\delta}_p(L) = \sum_{k=1}^{p-1} \bar{\delta}_{p,k} L^{k-1}$ and $\bar{\delta}_{p,k} = \sum_{n=k+1}^{p} \theta_n$. This implies

$$[\alpha - \delta_p(L)(\alpha - 1)]y_{t-1} = [\alpha - \delta_p(1)(\alpha - 1)]y_{t-1} - [\delta_p(L) - \delta_p(1)](\alpha - 1)y_{t-1}$$
$$= [\alpha - \delta_p(1)(\alpha - 1)]y_{t-1} + (\alpha - 1)\bar{\delta}_p(L)\Delta y_{t-1}. \tag{11}$$

Hence, by collecting the terms,

$$y_t = \alpha y_{t-1} + \delta_p(L)\Delta y_{t-1} - \delta_p(L)(\alpha - 1)y_{t-2} + \varepsilon_{p,t}$$
$$= [\alpha - \delta_p(L)(\alpha - 1)]y_{t-1} + \alpha\delta_p(L)\Delta y_{t-1} + \varepsilon_{p,t}$$
$$= [\alpha - \delta_p(1)(\alpha - 1)]y_{t-1} + [\alpha\delta_p(L) + (\alpha - 1)\bar{\delta}_p(L)]\Delta y_{t-1} + \varepsilon_{p,t}, \tag{12}$$

which is (9) with

$$\beta = \alpha - \delta_p(1)(\alpha - 1), \tag{13}$$
$$\beta_p(L) = \alpha\delta_p(L) + (\alpha - 1)\bar{\delta}_p(L), \tag{14}$$
$$e_{p,t} = \varepsilon_{p,t}. \tag{15}$$

This is important, for (at least) two reasons. One reason is that it shows how unless $\alpha = 1$ ($c = 0$), such that $\beta = \alpha$, α is not identified. This means that in the regression to be estimated the drift away from a unit root is not determined by c alone, but is in fact affected also by $\delta_p(1)$, as is clear from

$$\beta = 1 + [1 - \delta_p(1)]cT^{-1}. \tag{16}$$

This has implications for studies such as Moon andPhillips (2000) and Phillips et al. (2001), where the purpose is to estimate c. Another reason for why the above result is important is that it shows how the regression error in (9) is exactly the same as under the unit root null. This is very convenient in that once the model has been reparameterized

as in (9), most of the main results regarding the accuracy of the approximation can be taken more or less directly form Chang and Park (2002). However, this requires $p \to \infty$. It is therefore convenient to treat p as a function T.

Assumption 4 $pT^{-1/2} \to 0$ *as* $p, T \to \infty$.

Assumption 4 restricts the rate at which p is allowed to increase with T, but is weak enough to enable lag selection by standard information criteria, such as AIC and BIC.

3 The ADF test statistic and its local asymptotic distribution

Let

$$A_T = \sum_{t=1}^{T} y_{t-1}\varepsilon_{p,t} - \left(\sum_{t=1}^{T} y_{t-1}x'_{p,t}\right)\left(\sum_{t=1}^{T} x_{p,t}x'_{p,t}\right)^{-1}\left(\sum_{t=1}^{T} x_{p,t}\varepsilon_{p,t}\right) \quad (17)$$

$$B_T = \sum_{t=1}^{T} y_{t-1}^2 - \left(\sum_{t=1}^{T} y_{t-1}x'_{p,t}\right)\left(\sum_{t=1}^{T} x_{p,t}x'_{p,t}\right)^{-1}\left(\sum_{t=1}^{T} x_{p,t}y_{t-1}\right) \quad (18)$$

$$C_T = \sum_{t=1}^{T} \varepsilon_{p,t}^2 - \left(\sum_{t=1}^{T} \varepsilon_{p,t}x'_{p,t}\right)\left(\sum_{t=1}^{T} x_{p,t}x'_{p,t}\right)^{-1}\left(\sum_{t=1}^{T} x_{p,t}\varepsilon_{p,t}\right), \quad (19)$$

where $x_{p,t} = (\Delta y_{t-1}, ..., \Delta y_{t-p})'$. It is important to remember that the OLS estimator of the coefficient of y_{t-1} in (9) is not really estimating α, but rather β. Let us therefore consider OLS estimator $\hat{\beta}$ of β and its standard error, which are such that

$$\hat{\beta} = \beta + A_T B_T^{-1}, \quad (20)$$

$$s(\hat{\beta})^2 = \hat{\sigma}^2 B_T^{-1}, \quad (21)$$

where $\hat{\sigma}^2 = T^{-1}(C_T - A_T^2 B_T^{-1})$. The test statistic of interest is the usual ADF statistic, which is given by

$$ADF = \frac{\hat{\beta} - 1}{s(\hat{\beta})}. \quad (22)$$

Lemmas 1 and 2, which are analogous to Lemmas 3.1 and 3.2 of Chang and Park (2002), are key in deriving the local asymptotic distribution of ADF.

Lemma 1 *Under Assumptions 1–3,*

(a) $\displaystyle T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_{p,t} = \pi(1)T^{-1}\sum_{t=1}^{T} w_{t-1}\varepsilon_t + o_p(1);$

(b) $\displaystyle T^{-2}\sum_{t=1}^{T} y_{t-1}^2 = \pi(1)^2 T^{-2}\sum_{t=1}^{T} w_{t-1}^2 + o_p(1);$

(c) $\quad T^{-1} \sum_{t=1}^{T} \varepsilon_{p,t}^2 = T^{-1} \sum_{t=1}^{T} \varepsilon_t^2 + o_p(p^{-1}),$

where $w_t = \sum_{n=1}^{t} \alpha^{t-n} \varepsilon_n$.

Lemma 2 *Under the conditions of Lemma 1,*

(a) $\quad \left\| \left(T^{-1} \sum_{t=1}^{T} x_{p,t} x_{p,t}' \right)^{-1} \right\| = O_p(1);$

(b) $\quad \left\| T^{-1} \sum_{t=1}^{T} x_{p,t} y_{t-1} \right\| = O_p(\sqrt{p});$

(c) $\quad \left\| T^{-1} \sum_{t=1}^{T} x_{p,t} \varepsilon_{p,t} \right\| = o_p(p^{-1/2}).$

The proofs of Lemmas 1 and 2 are almost identical to the proofs of Lemmas 3.1 and 3.2 in Chang and Park (2002), and are therefore omitted. The only difference is the presence of α in w_t, which does not affect the derivations.[2] Lemmas 1 and 2 imply that

$$T^{-1} A_T = \pi(1) T^{-1} \sum_{t=1}^{T} w_{t-1} \varepsilon_t + o_p(1) \tag{23}$$

$$T^{-2} B_T = \pi(1)^2 T^{-2} \sum_{t=1}^{T} w_{t-1}^2 + O_p(pT^{-1}) \tag{24}$$

$$T^{-1} C_T = T^{-1} \sum_{t=1}^{T} \varepsilon_{p,t}^2 + o_p(p^{-1}), \tag{25}$$

where the remainder terms are all $o_p(1)$ under Assumption 4. In view of Lemma 1 (c), this implies

$$\hat{\sigma}^2 = T^{-1}(C_T - A_T^2 B_T^{-1}) = T^{-1} C_T - T^{-1}(T^{-1} A_T)^2 (T^{-2} B_T)^{-1} = T^{-1} C_T + o_p(1)$$

$$= T^{-1} \sum_{t=1}^{T} \varepsilon_t^2 + o_p(1) \rightarrow_p \sigma^2 \tag{26}$$

---

[2] Some of the orders reported in Lemmas 1 and 2 are not sharp. For example, as pointed out by Chang and Park (2002), under Assumption 2 the remainder in Lemma 1 (c) can be reduced to $o_p(p^{-s})$.

(see Chang and Park 2002, Proof of Lemma 3.3). Let us now consider ADF. Note how $\beta - 1 = c[1 - \delta_p(1)]T^{-1}$. Together with Lemmas 1 and 2, this implies

$$
\begin{aligned}
ADF &= \frac{\hat{\beta} - \beta}{s(\hat{\beta})} + \frac{\beta - 1}{s(\hat{\beta})} \\
&= \hat{\sigma}^{-1} \left[T^{-1} A_T (T^{-2} B_T)^{-1/2} + c[1 - \delta_p(1)]\sqrt{T^{-2} B_T} \right] \\
&= \sigma^{-1} \left[\frac{T^{-1} \sum_{t=1}^{T} w_{t-1}\varepsilon_t}{\sqrt{T^{-2} \sum_{t=1}^{T} w_{t-1}^2}} + c[1-\delta_p(1)]\pi(1) \left(T^{-2} \sum_{t=1}^{T} w_{t-1}^2 \right)^{1/2} \right] + o_p(1).
\end{aligned}
\tag{27}
$$

The asymptotic distribution of the right-hand side is easily evaluated using the results provided in Hansen (1995) for the finite-order AR case, and is summarized in Theorem 1.

Theorem 1 *Under Assumptions 1–4,*

$$
ADF \to_w \frac{\int_{r=0}^{1} J_c(r)dW(r)}{\sqrt{\int_{r=0}^{1} J_c(r)^2 dr}} + c \lim_{p \to \infty} [1 - \delta_p(1)]\pi(1) \cdot \left(\int_{r=0}^{1} J_c(r)^2 dr \right)^{1/2},
$$

where $J_c(r) = \int_{v=0}^{r} \exp[c(r - v)]dW(v)$ *with* $W(r)$ *being a standard Brownian motion on* $r \in [0, 1]$.

Phillips (1987) considers the (non-augmented) Dickey–Fuller test statistic in the case of serially uncorrelated errors. The difference between the local asymptotic distribution reported in Theorem 1 and the one given in Phillips (1987) is the presence of $[1 - \delta_p(1)]\pi(1)$. It is therefore interesting to consider briefly the behaviour of this term. Note how $\theta(1) = 1 - \delta(1)$, which implies $[1 - \delta_p(1)] \to \theta(1)$ as $p \to \infty$. But $\theta(1) = \pi(1)^{-1}$, and so

$$
\lim_{p \to \infty} [1 - \delta_p(1)]\pi(1) = 1.
\tag{28}
$$

The effect of the truncation on the asymptotic distribution of the ADF test statistic is therefore negligible. This finding is in stark contrast to the results reported by Ng and Perron (2001) and Paparoditis and Politis (2017), where the effect of p is non-negligible. In practice, of course, p is fixed, which means that $[1 - \delta_p(1)]\pi(1) \neq 1$. The asymptotic null distribution of ADF under $c = 0$ is given by

$$
ADF \to_w \frac{\int_{r=0}^{1} W(r)dW(r)}{\sqrt{\int_{r=0}^{1} W(r)^2 dr}},
\tag{29}
$$

which is independent of $[1 - \delta_p(1)]\pi(1)$. One of the effects of the truncation is therefore to affect the drift of the distribution under the alternative hypothesis that

$c < 0$. Hence, while negligible, in finite samples we expect p to have an effect on power. This prediction is in agreement with the bulk of the existing Monte Carlo evidence (see, for example, Ng and Perron 1995). In fact, the local power predictions derived here seem very accurate, even when compared to the stationary predictions of Paparoditis and Politis (2017) when the data are generated as stationary. Let us explain what we mean by this. Paparoditis and Politis (2017) show that the power of the ADF test against stationary alternatives should be decreasing in p, even asymptotically. This is their theoretical prediction. They then simulate power under $\alpha \in \{0.985, 0.97\}$, $\pi(L) = 1 + \pi_1 L$, $\pi_1 \in \{-0.5, 0.5\}$, $T \in \{50, 100, 200, 400, 800, 1600\}$ and $p = T^a$ with a going from 0.05 to 0.49 in steps of 0.04. Except for the non-local specification of α, this is consistent with the DGP considered here. Note in particular how p satisfies our Assumption 4. According to the results reported in their Table 6 for the case when $\alpha = 0.97$ and $\pi_1 = -0.5$ (in which the effect of p is most pronounced), while when $T = 50$ power decreases almost monotonically from 0.17 when $a = 0.05$ to 0.09 when $a = 0.49$, when $T = 1600$ power is flat at 1. Clearly, this finding does not fit well with the prediction that power should always decrease with increases in p. It is, however, consistent with our prediction that the effect of p should tend to decrease with increasing T.

Remark 2 As already mentioned, Chang and Park (2002) only consider the asymptotic distribution under the unit root null. They also claim (without proof) in their Remark 3.2 that the asymptotic distribution under Assumption 3 with $c \neq 0$ should be the same, but with $W(r)$ replaced by $J_c(r)$. In order to asset the validity of this claim, note how $dJ_c(r) = cJ_c(r)dr + W(r)$, implying

$$\frac{\int_{r=0}^{1} J_c(r)dJ_c(r)}{\sqrt{\int_{r=0}^{1} J_c(r)^2 dr}} = \frac{\int_{r=0}^{1} J_c(r)dW(r)}{\sqrt{\int_{r=0}^{1} J_c(r)^2 dr}} + c \left(\int_{r=0}^{1} J_c(r)^2 dr \right)^{1/2}, \qquad (30)$$

which is identically the local asymptotic distribution reported by Phillips (1987). The fact that this distribution is also the limit of the local asymptotic distribution in Theorem 1 as $p \to \infty$ proves that the claim of Chang and Park (2002) is in fact correct.

Remark 3 As discussed in Remark 3.1 of Chang and Park (2002), DGPs with deterministic constant and trend terms can be easily accommodated. Such an extension is interesting not only in its own right, but also because it shows how the results reported here extends to other unit root tests. Let us therefore use z_t to denote the observed data. A common way to accommodate deterministic constant and trend terms is through the following components model: $z_t = \mu + \tau t + y_t$, where y_t is as in (1). In this DGP, testing for a unit root in z_t is equivalent to testing for a unit root in y_t. The problem is how to purge the effect of the deterministic terms. Chang and Park (2002) discuss the case when this is done through an auxiliary OLS regression of z_t onto a constant or a constant and trend. In this case, the results reported in this paper are the same, except that $J_c(r)$ has to be replaced by its suitably demeaned or detrended version, $J_c^d(r)$ say. Specifically, while in the constant-only case case, $J_c^d(r) = J_c(r) - \int_{v=0}^{1} J_c(v)dv$, in the case with both a constant and trend,

$J_c^d(r) = J_c(r) + (6r - 4) \int_{v=0}^1 J_c(v)dv - (12r - 6) \int_{v=0}^1 v J_c(v)dv$. An alternative to OLS is to perform generalized least squares (GLS) under the local alternative, as first suggested by Elliott et al. (1996). As Westerlund (2014) shows, except for $[1 - \delta_p(1)]\pi(1)$, the asymptotic distribution of the resulting ADF–GLS test in the constant-only case is identical to the one given in Theorem 1. The results reported here regarding the effect of p therefore apply also this other test. Another possibility is to follow, for example, Shin and So (2001) and to perform the OLS demeaning recursively. The asymptotic distribution in this case is again the same as in Theorem 1 but now with $J_c(r)$ replaced by $J_c^d(r) = J_c(r) - r^{-1} \int_{v=0}^r J_c(v)dv$. The asymptotic distributions of these other tests in the trend case do not have the same form as in Theorem 1, but the effect of p is still expected to be negligible. Moreover, these results extend quite naturally to the bulk of the existing panel data unit root tests, which are typically nothing but panel extensions of known time series tests (see, for example, Westerlund 2016, for a discussion of the issue of parametric lag correction in the panel data context).

References

Chang Y, Park JY (2002) On the asymptotics of ADF tests for unit roots. Econom Rev 21:431–447

Elliott G, Rothenberg TJ, Stock JH (1996) Efficient tests for an autoregressive unit root. Econometrica 64:813–836

Hansen BE (1995) Rethinking the univariate approach to unit root testing. Econom Theory 11:1148–1171

Lopez JH (1997) The power of the adf test. Econom Lett 57:5–10

Moon HR, Phillips PC (2000) Estimation of autoregressive roots near unity using panel data. Econom Theory 16:927–997

Ng S, Perron P (1995) Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. J Am Stat Assoc 90:268–281

Ng S, Perron P (2001) Lag length selection and the construction of unit root tests with good size and power. Econometrica 69:1519–1554

Paparoditis E, Politis DN (2017) The asymptotic size and power of the augmented dickey-fuller test for a unit root. Econom Rev (forthcoming)

Phillips PC (1987) Towards a unified asymptotic theory for autoregression. Biometrika 74:535–547

Phillips PC, Moon HR, Xiao Z (2001) How to estimate autoregressive roots near unity. Econom Theory 17:29–69

Phillips PC, Solo V (1992) Asymptotics for linear processes. Ann Stat 971–1001

Said SE, Dickey DA (1984) Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika 71:599–607

Shin DW, So BS (2001) Recursive mean adjustment for unit root tests. J Time Ser Anal 22:595–612

Stock JH (1991) Confidence intervals for the largest autoregressive root in US macroeconomic time series. J Monet Econ 28:435–459

Westerlund J (2014) On the asymptotic distribution of the Dickey Fuller-GLS test statistic. Statistics 48:1233–1253

Westerlund J (2016) The asymptotic distribution of the CADF unit root test in the presence of heterogeneous AR(p) errors. Stat Pap 57:303–317

Xiao Z, Phillips PCB (1998) An ADF coefficient test for a unit root in ARMA models of unknown order with empirical applications to the us economy. Econom J 1:27–43