

## Article

# Prepivoted Augmented Dickey-Fuller Test with Bootstrap-Assisted Lag Length Selection

Somak Maitra <sup>1</sup> and Dimitris N. Politis <sup>2,\*</sup>

<sup>1</sup> Department of Mathematics, University of California—San Diego, La Jolla, CA 92093, USA; somaitra@ucsd.edu

<sup>2</sup> Department of Mathematics and Halıcıoğlu Data Science Institute, University of California—San Diego, La Jolla, CA 92093, USA

\* Correspondence: dpolitis@ucsd.edu

**Abstract:** We investigate the application of prepivoting in conjunction with lag length selection to correct the size and power performance of the Augmented Dickey-Fuller test for a unit root. The bootstrap methodology used to perform the prepivoting is a residual based AR bootstrap that ensures that bootstrap replicate time series are created under the null irrespective of whether the originally observed series obeys the null hypothesis or not. Simulation studies wherein we examine the performance of our proposed method are given; we evaluate our method’s performance on ARMA(1,1) models with varying configurations for size and power performance. We also propose a novel data dependent lag selection technique that uses bootstrap data under the null to select an optimal lag length; the performance of our method is compared to existing lag length selection criteria.

**Keywords:** ADF test; bootstrap; hypothesis testing; unit root; ARMA models; lag length selection



**Citation:** Maitra, S.; Politis, D.N. Prepivoted Augmented Dickey-Fuller Test with Bootstrap-Assisted Lag Length Selection. *Stats* **2024**, *7*, 1226–1243. <https://doi.org/10.3390/stats7040072>

Academic Editors: Marco André da Silva Costa and Magda Sofia Valéria Monteiro

Received: 25 August 2024

Revised: 1 October 2024

Accepted: 9 October 2024

Published: 17 October 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Unit root testing is a central problem of time series and econometric analysis. For example, the presence (or not) of a unit root in time series data has huge repercussions in forecasting applications. In econometric analyses, unit roots render the data to be highly sensitive to shocks; while stationary processes tend to revert to the mean in the presence of shocks, random walks tend to diverge.

The primary approaches to unit root testing include the augmented Dickey-Fuller (henceforth referred to as ADF) test [1], and the Phillips-Perron (PP) test [2]. Extensive literature including the work of Schwert (1989) [3], Perron and Ng (1996) [4] as well as Davidson and MacKinnon (2004) [5] point to evidence that PP tests underperform with respect to size in finite samples compared to ADF tests since PP tests are highly reliant on asymptotic results; additionally PP tests suffer from severe size distortions in the presence of negative moving average errors. However, the ADF test may suffer from reduced power; see Paparoditis and Politis (2018) [6] and references therein.

The ADF test examines the presence of a unit root in a stretch of time series data  $X_1, X_2, \dots, X_n$  by focusing on the Ordinary Least Squares (OLS) estimate  $\hat{\rho}$  of  $\rho$  in the regression equation

$$X_t = \rho X_{t-1} + \sum_{i=1}^q a_{j,q} \Delta X_{t-i} + \epsilon_{t,q} \quad (1)$$

fitted to the observed data. In the above equation, we use the notation  $\Delta X_t = X_t - X_{t-1}$ . Additionally, note that the number of lagged differences (denoted by  $q$ ) is allowed to vary with the sample size  $n$ , and thus  $q$  is an abbreviated notation for  $q_n$ . The null and alternative hypotheses for the ADF testing framework are as follows.

$$H_0 : X_t = X_{t-1} + U_t \quad (2)$$

$$H_a : X_t = U_t \quad (3)$$

A typical assumption in the above is that  $U_t = \sum_{j=1}^{\infty} a_j U_{t-j} + e_t$ , where the  $a_j$  are autoregressive (AR) coefficients, and  $e_t$  is a sequence of independent, identically distributed (i.i.d.) mean zero random variables with finite (and nonzero) variance, denoted by  $\sigma_e^2$ ; however, the i.i.d. assumption on  $e_t$  can be relaxed as discussed in Section 2. To elaborate, the hypotheses are essentially that under the null, the series  $X_t$  is obtained by integrating an infinite order autoregressive process, while the alternative is that the series is a *de facto* infinite order autoregressive process. The process  $U_t$  can be guaranteed to be stationary and causal by imposing conditions on its coefficients, namely  $\sum_{j=1}^{\infty} |j|^s |a_j| < \infty$  for some  $s \geq 1$  and that  $\sum_{j=1}^{\infty} a_j z^j \neq 0$  for all  $|z| \leq 1$ .

In their seminal work on the topic, Dickey and Fuller (1979) [1] suggested the studentized statistic

$$t_{n,q} = \frac{\hat{\rho}_n - 1}{\widehat{\text{Std}}(\hat{\rho}_n)}$$

with  $\widehat{\text{Std}}(\hat{\rho}_n)$  used to denote the estimated standard deviation of the OLS estimator of  $\rho$ . We explicitly denote the importance of the lag length  $q$  by using the notation  $t_{n,q}$  to denote the ADF test statistic as opposed to  $t_n$ . The distribution of this studentized statistic under the null is non-standard/non-normal; however, it is free of unknown parameters, and has been promulgated in econometric and statistics literature. Assumptions under which this distribution is valid have been progressively relaxed since Dickey and Fuller's original work in 1979; see [6] for a chronological account.

The ADF test has been demonstrated to have less than ideal size and power performance in real world applications. It is well examined in literature [3,7] that the presence of negative MA coefficients in the data generating process causes severe size distortions, while the work of Paparoditis and Politis [6] offers a concrete real world example wherein the power performance is poor. The latter work also provides evidence of the asymptotic collinearity problem in the ADF regression; the collinearity becomes more prominent for large  $q$ , in which case loss of power ensues. The issue is then how to work with a moderate value of  $q$  while still achieving a (close to) nominal size of the ADF test.

A succinct summary justifying the choice of adopting the pre pivoting framework for the ADF test comes from Section 3 of Beran (1988) [8], where the author demonstrates that (under regularity conditions and in the case where the asymptotic null distribution of the test statistic is independent of the unknown parameters) pre pivoted tests have smaller order errors in rejection probability than that of the asymptotic theory test. Since the asymptotic null distribution of the OLS estimate  $\hat{\rho}_n$  from the ADF regression is independent of  $\rho$ , the ADF test belongs to the category of hypothesis tests that stand to benefit from the pre pivoting framework.

The remainder of this paper is organized as follows. Section 2 reviews the asymptotic properties of the ADF test statistic, along with conditions on  $q_n$  for good size and power. We also demonstrate that the AR-sieve bootstrap of Paparoditis and Politis [9] is a pre pivoted test in the sense of Beran [8]. Section 3 highlights the consistency of the pre pivoted ADF test against fixed and local alternatives. Section 4 details the results of numerical experiments wherein we study the empirical performance of the pre pivoted ADF test for different lag length specifications. Section 5 introduces a novel bootstrap-based tuning parameter selection algorithm and its application to the pre pivoted ADF test, along with numerical experiments that demonstrate its efficacy with respect to both size and power.

## 2. Asymptotic Properties of the ADF Test

The asymptotic properties of the ADF test primarily hinge upon the convergence results of the test statistic  $t_{n,q}$  under the null and the alternative, as well as the conditions governing the underlying stationary process  $U_t$ . In what follows, we proffer a brief summary of the relaxations to assumptions under which the convergence result of the test

statistic  $t_{n,q}$  under the null. Dickey and Fuller [1] initially derived the non-standard null distribution of the statistic under the assumption that the underlying process is autoregressive with *known* and finite order. The result was consequently extended in 1984 by Said and Dickey [10] to the setup where the innovation process  $U_t$  is an invertible ARMA process; the latter can be expressed as an AR( $\infty$ ) process with exponentially decaying coefficients. Further relaxations to conditions have been examined in literature with Chang and Park (2002) [11] considering the case that  $\{e_t\}$  is a martingale difference sequence and in 2018, Paparoditis and Politis [6] showed the convergence under the sole provision that the innovation process  $U_t$  has a continuous spectral density that is strictly positive; this is can also be stated in the equivalent form that the process  $U_t$  has a Wold-type AR representation with respect to just white noise errors, allowing for a much larger class of time series versus hitherto linear AR( $\infty$ ) processes with i.i.d. or martingale difference innovations. The following subsections review the two main asymptotic results of the ADF test statistic from the work of Paparoditis and Politis (2018) [6].

### 2.1. Behaviour of the Test Statistic Under the Null

We now describe the convergence behavior of the test statistic  $t_{n,q}$ . We adopt the general framework of Paparoditis and Politis [6], and assume that  $U_t$  is a mean zero weakly stationary process with autocovariance function  $\gamma_U(h) = \text{Cov}(U_0, U_h)$  that is absolutely summable; hence, the spectral density  $f_U(w) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \gamma_U(h) e^{iwh}$  is well-defined. We also assume that the logarithm of the spectral density  $f_U$  is integrable, and therefore  $U_t$  admits the Wold representation

$$U_t = \sum_{j=1}^{\infty} \alpha_j \epsilon_{t-j} + \epsilon_t$$

where  $\sum_{j=1}^{\infty} \alpha_j^2 < \infty$  and  $\epsilon_t$  is a white noise, i.e., a weakly stationary process with zero mean, common variance and  $\text{Cov}(\epsilon_t, \epsilon_k) = 0$  for  $t \neq k$ ; see Brockwell and Davis [12].

If we further assume that  $f_U(w) > 0$  for all  $w$ , then  $U_t$  also admits the Wold-type AR representation:

$$U_t = \sum_{j=1}^{\infty} b_j U_{t-j} + \epsilon_t \quad (4)$$

where  $\epsilon_t$  is the same white noise process as above, and the coefficients  $b_j$  are absolutely summable. Additionally,  $b(z) = 1 - \sum_{j=1}^{\infty} b_j z^j \neq 0$  for  $|z| \leq 1$ ; see [6] for details. The assumption of positive spectral density of the underlying innovation process is a sine qua non for an AR( $\infty$ ) approximation to the spectral density to be consistent; this assumption has been used by Paparoditis and Politis (2018) [6] and Kreiss, Paparoditis and Politis (2011) [13] and has not been found to be a limitation for practical applications and has not been found to be a limitation for practical applications.

Under all the above assumptions, and recalling the ADF regression (1) used to test unit root hypothesis  $H_0$ , the following result is true. To state it, we require an extra assumption on the white noise process appearing in (4), namely:

$$\mathbb{E}(\epsilon_t^4) < \infty \text{ and } \sum_{n=1}^{\infty} n \|\mathcal{P}_1(\epsilon_n)\| < \infty \quad (5)$$

where  $\mathcal{P}_t(Y) = \mathbb{E}[Y|\mathcal{F}_t] - \mathbb{E}[Y|\mathcal{F}_{t-1}]$ ,  $\mathcal{F}_s$  is the  $\sigma$ -field generated by  $(\epsilon_t \text{ for } t \leq s)$ , and the norm  $\|\cdot\|$  is the  $L_p$  norm with  $p = 4$ .

**Theorem 1** (Paparoditis and Politis (2018) [6]). Assume  $f_U(w) > 0$  for all  $w$ , and Equations (4) and (5). If  $p_n \rightarrow \infty$  as  $n \rightarrow \infty$  such that  $\frac{p_n}{\sqrt{n}} \rightarrow 0$ , then

$$t_{n,q} \xrightarrow{d} \frac{\int_0^1 W(t)dW(t)}{(\int_0^1 (W(t))^2 dt)^{\frac{1}{2}}}$$

when  $H_0$  is true; here  $W(t)$ ,  $t \in [0, 1]$  is the standard Wiener process on  $[0, 1]$ , and  $\xrightarrow{d}$  denotes convergence in distribution.

The  $\alpha$  level ADF test therefore rejects the null  $H_0$  whenever the test statistic  $t_{n,q}$  is smaller than  $C_\alpha$  where  $C_\alpha$  is the lower  $\alpha$  percentile of the above distribution.

Theorem 1 can be further extended to the cases wherein (1) is modified to include an affine time trend; the same limit result except that the standard Wiener process is replaced by

$$\tilde{W}_t = W_t + (6t - 4) \int_0^1 W(s)ds - (12t - 6) \int_0^1 sW(s)ds$$

For a case by case analysis of intercept/trend inclusion, the reader is referred to the book of Hamilton [14].

## 2.2. Behaviour Under the Alternative

The limiting distribution of the test statistic  $t_{n,q}$  under the alternative is more straightforward; the limiting distribution is normal albeit dependent on the true  $\rho$  in (1). The work of Paparoditis and Politis [6] establishes this limiting behavior even under their relaxed requirement where the underlying innovation process is just assumed to possess a continuous spectral density which is strictly positive; they also discuss the problem of asymptotic collinearity in the ADF regression where increasing the number of lagged differences leads to a reduction in power. In essence, as the chosen number of lags  $q$  increases, the regressors in the ADF regression (1) become asymptotically collinear, leading to slow rate of convergence of the estimator  $\hat{\rho}_n$ . This directly leads to a loss of power of the ADF test. We comment on this behaviour of the prepivoted ADF test further in the following sections.

Under the same assumptions on the innovation process  $U_t$  used in Theorem 1 the limiting behaviour of the test statistic under the alternative hypothesis  $H_a$  is as follows.

**Theorem 2** (Paparoditis and Politis (2018) [6]). Assume  $X_t = U_t$  is true with  $U_t$  such that  $f_U(w) > 0$  for all  $w$ , and Equations (4) and (5). Let  $q = q_n \rightarrow \infty$  as  $n \rightarrow \infty$  in such a way that  $\frac{q_n^{\frac{1}{2}}}{n} \rightarrow 0$  and  $\sqrt{n} \sum_{j=q+1}^{\infty} |a_j| \rightarrow 0$ . Then, as  $n \rightarrow \infty$

1.  $\frac{n}{q} \text{Var}(\hat{\rho}_n) \rightarrow (1 - \rho)^2$  in probability
2.  $\sqrt{\frac{n}{q}}(\hat{\rho}_n - \rho) \xrightarrow{d} \mathcal{N}(0, (1 - \rho)^2)$ .

## 3. Power of the Prepivoted Test

Prepivoting, as introduced by Beran [8], is the mapping of the test statistic  $t_{n,q}$  to a new test statistic  $F^*(t_{n,q})$ , where  $F^*(\cdot)$  is a consistent estimator of the distribution function of  $t_{n,q}$  computed from the data  $(X_1, \dots, X_n)$ ; typically,  $F^*(\cdot)$  will be based on some kind of bootstrap procedure. We will adopt the residual AR bootstrap of Paparoditis and Politis [9] to generate the  $i$ th bootstrap sample denoted by  $(X_{1,i}^*, \dots, X_{n,i}^*)$  where  $i = 1, \dots, B$ . Applying the ADF regression (1) to each of these  $B$  samples, we obtain  $B$  bootstrap ADF statistics  $t_{n,q}^*$ ; this allows us to compute the bootstrap estimate of the CDF, denoted by  $F^*(\cdot)$ . Thus, the prepivoted test statistic is  $F^*(t_{n,q}) = \frac{1}{B} \sum_{i=1}^B \mathbb{1}_{\{t_{n,q}^* < t_{n,q}\}}$ . The prepivoted ADF test rejects the null hypothesis if  $F^*(t_{n,q}) < \alpha$ , the nominal level of the test.

In this section, we briefly discuss the power of the prepivot ADF test. In particular, we show that it is consistent against the alternative hypothesis  $\rho < 1$ . We build upon the result of Theorem 2 and obtain the following novel result.

**Theorem 3.** *If  $q = q_n \rightarrow \infty$  as  $n \rightarrow \infty$  such that  $\frac{q_n^4}{n} \rightarrow 0$  and  $\sqrt{n} \sum_{j=q+1}^{\infty} |a_j| \rightarrow 0$ , then the prepivot ADF test is consistent where the alternative is that  $\rho$  is any fixed value less than 1.*

**Proof.** From Theorem 2 we have that  $\sqrt{\frac{n}{q}}(\hat{\rho}_n - \rho) \xrightarrow{d} \mathcal{N}(0, (1 - \rho)^2)$ . Observe that the ADF test statistic can be written as

$$t_{n,q} = \frac{\hat{\rho}_n - 1}{\widehat{\text{Std}}(\hat{\rho}_n)} = \frac{\hat{\rho}_n - \rho + \rho - 1}{\widehat{\text{Std}}(\hat{\rho}_n)}$$

The power of the prepivot test is given by the probability of correct rejection of the test, i.e.,  $\mathbb{P}(F^*(t_{n,q}) < \alpha | H_a)$  where  $F^*$  is the bootstrap estimate of the CDF of  $t_{n,q}$ . Let  $C_\alpha^*$  be an  $\alpha$  quantile of  $F^*$ .

Then,

$$\begin{aligned} \mathbb{P}(F^*(t_{n,q}) < \alpha | H_a) &= \mathbb{P}(t_{n,q} < C_\alpha^* | H_a) \\ &= \mathbb{P}((\hat{\rho}_n - 1) < C_\alpha^* \widehat{\text{Std}}(\hat{\rho}_n) | H_a) \\ &= \mathbb{P}(\hat{\rho}_n - \rho < C_\alpha^* \widehat{\text{Std}}(\hat{\rho}_n) + (1 - \rho) | H_a) \\ &= \mathbb{P}\left(\sqrt{\frac{n}{q}} \frac{\hat{\rho}_n - \rho}{1 - \rho} < C_\alpha^* \sqrt{\frac{n}{q}} \frac{\widehat{\text{Std}}(\hat{\rho}_n)}{1 - \rho} + \sqrt{\frac{n}{q}} \Big| H_a\right) \\ &= \Phi\left(C_\alpha^* \sqrt{\frac{n}{q}} \frac{\widehat{\text{Std}}(\hat{\rho}_n)}{1 - \rho} + \sqrt{\frac{n}{q}}\right) \end{aligned}$$

Now since  $\widehat{\text{Std}}(\hat{\rho}_n) = \sqrt{\frac{q}{n}}(1 - \rho) + o_p(\sqrt{\frac{q}{n}})$  and using consistency of bootstrap quantiles (cf. Lemma 1.2.1 of Politis, Romano and Wolf [15]) and the continuous mapping theorem, we have

$$\Phi\left(C_\alpha^* \sqrt{\frac{n}{q}} \frac{\widehat{\text{Std}}(\hat{\rho}_n)}{1 - \rho} + \sqrt{\frac{n}{q}}\right) \approx \Phi\left(C_\alpha + \sqrt{\frac{n}{q}}\right) \quad (6)$$

where  $C_\alpha$  is the  $\alpha$  quantile of the Dickey-Fuller distribution. Observe therefore, that the power of the prepivot test tends to 1 as  $n \rightarrow \infty$ .

While it is not immediately clear from this expression, the prepivot ADF test in fact has better power than the asymptotic ADF test. We demonstrate the same through numerical experiments in the next two sections.  $\square$

We briefly discuss the behavior of the prepivot ADF test in the local alternative framework. We build upon the theorem of Aylar, Westerlund and Smeekes [16], where they derive the limiting distribution of the ADF test statistic  $t_{n,q}$  under the local alternative  $\rho = 1 + \frac{c}{n}$  for some fixed  $c < 0$  to demonstrate non-trivial power of the prepivot ADF test. Their notation and assumptions are briefly reviewed as follows. The DGP under consideration is  $X_t = \rho X_{t-1} + U_t$  with  $U_t = \pi(L)\epsilon_t$ . It is assumed that  $\epsilon_t$  is a martingale difference sequence with some filtration  $\mathcal{F}_t$  with  $E(\epsilon_t^2) = \sigma^2$ ,  $\frac{1}{n} \sum_{t=1}^n \epsilon_t^2 \rightarrow_p \sigma^2$  and  $E(|\epsilon_t^4|) < \infty$ . Assume  $\pi(z) \neq 0$  for all  $|z| \leq 1$ , and  $\sum_{k=0}^{\infty} |k|^s |\pi_k| < \infty$  for some  $s \geq 1$ . They then assume  $\rho = 1 + \frac{c}{n}$  with  $c < 0$ , and  $\frac{q}{\sqrt{n}} \rightarrow 0$  as  $n \rightarrow \infty$ . Then, they claim that

$$t_{n,q} \xrightarrow{d} \frac{\int_0^1 J_c(r) dW(r)}{\left(\int_0^1 J_c^2(r) dr\right)^{\frac{1}{2}}} + c \lim_{q \rightarrow \infty} (1 - \delta_q(1)) \pi(1) \left( \int_0^1 J_c^2(r) dr \right)^{\frac{1}{2}} \quad (7)$$

In the above Equation (7),  $J_c(r) = \int_{v=0}^r \exp(c(r-v))dW(v)$  where  $W(r)$  is the standard Wiener process on  $[0, 1]$ . Additionally,  $\delta_q(L) = \sum_{k=1}^q \theta_k L^{k-1}$ , where  $\theta(L) = \pi(L)^{-1} = 1 - \sum_{k=1}^{\infty} \theta_k L^k$ . Given this setup, we have the following result.

**Theorem 4.** *Under the above assumptions, the prepivot ADF test has non-trivial power against a local alternative of the form  $\rho = 1 + \frac{c}{n}$  where  $c$  is a negative constant.*

**Proof.** Let  $G(\cdot)$  be the continuous CDF of the limiting distribution given in (7). Note that  $J_c(r) \sim \mathcal{N}(0, \frac{e^{2rc}-1}{2c})$  for  $r \in [0, 1]$  and is an Ornstein-Uhlenbeck process that solves  $dJ_c(r) = cJ_c(r) + dW(r)$  with  $J_c(0) = 0$ . Let  $H_{\text{loc}}$  denote the local alternative hypothesis. Then,

$$\begin{aligned} \mathbb{P}(F^*(t_{n,q}) < \alpha | H_{\text{loc}}) &= \mathbb{P}(t_{n,q} < C_{\alpha}^* | H_{\text{loc}}) \\ &\approx P(t_{n,q} < C_{\alpha} | H_{\text{loc}}) + O_p(C_{\alpha}^* - C_{\alpha}) \approx G(C_{\alpha}) \neq 0 \end{aligned}$$

where the last line follows since  $G(\cdot)$  has infinite support.  $\square$

**Remark 1.** *The above mentioned result of Aylar, Smeekes and Westerlund fails to capture the loss of power with increasing  $q$ . The asymptotic collinearity problem is empirically verified in Paparoditis and Politis [6] as well as the later sections of the paper at hand. Nevertheless, the result does allow us to demonstrate that the prepivot ADF test has non-trivial power against local alternatives—which can be seen in Sections 4 and 5, in particular Tables 3 and 4.*

#### 4. Numerical Experiments

In this section we provide the results of a numerical experiment. The setup is similar to that of Paparoditis and Politis [6], wherein the authors consider the ARMA(1,1) model to be the data generating process i.e.,  $X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}$  with  $Z_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ . We consider six different combinations of the ARMA parameters  $\phi$  and  $\theta$ ; two that yield samples under the unit root hypothesis and four that yield samples under the alternative hypothesis. For each sample size  $n$ , we generate 10,000 time series each of length  $n$ .

This simulation allowed us to compute the empirical rejection probabilities of the prepivot ADF test at the nominal level 0.05. We use a ‘tuning parameter sweep’ approach to select the lag lengths for the experiment in this section: we used the formula  $q = n^a$  with varying values of  $a$ . Note that the lag length used was the same for both the ADF regression on the original data as well as the ADF regression on the bootstrap samples.

The bootstrap methodology selected to generate bootstrap samples under the null is from Section 2.2 of Paparoditis and Politis [9]. This is a bootstrap approach that is based on unrestricted residuals, and the resultant bootstrap samples are generated under the null irrespective of whether the original data obey the null or not.

Tables 1 and 2 correspond to the setting where  $\{X_t\}$  has a unit root. The resulting empirical rejection probabilities should be close to the nominal level of the test i.e., 0.05. The entry closest to 0.05 is presented in bold face. As opposed to [6], the column  $a = 0.05$  is omitted in all the Tables—the entries in all 6 Tables corresponding to  $a = 0.05$  were very similar to those of  $a = 0.09$ .

From the results of Tables 1 and 2, we observe that the prepivot ADF test yields accurate size when the lag length is large. The test has good size in the presence of the negative MA parameter  $\theta = -0.5$  when lags are large, but suffers from severe over-rejection of the null when the lag lengths are short. The test is able to achieve accurate size for all sample sizes.

Tables 3–6 correspond to the setting where  $\{X_t\}$  does not have a unit root and is under the alternative hypothesis. These empirical rejection probabilities therefore represent the power of the prepivot ADF test, and should ideally be as large as possible. The largest entries in each row are represented in boldface.

The results of Tables 3–6 are encouraging. While the finite sample power of the prepivot ADF test is fairly low, it is much higher than the power of the asymptotic ADF

test for the same specifications. We observe that the power tends to 1 as the sample size increases—empirically verifying the theoretical consistency of the test. Unfortunately, we see that the lag length chosen to optimize size does not also optimize power—the power of the test appears to be highest when the chosen lag length is short. We also see the manifestation of the asymptotic collinearity problem from Paparoditis and Politis [6] when the lag lengths are allowed to increase, we observe that the power of the test reduces. In the next section, we will develop a bootstrap method with the goal of choosing a lag length that optimizes power while securing a size close to nominal.

**Table 1.** Entries represent empirical rejection probabilities of the prepivot ADF test with  $q = n^a$  for various values of  $a$ . The data generating process was ARMA(1,1) with  $(\phi, \theta) = (1, 0.5)$ .

$a =$	0.09	0.13	0.17	0.21	0.25	0.29	0.33	0.37	0.41	0.45	0.49
$n = 50$	0.0845	0.0916	0.0848	0.0380	0.0390	0.0585	0.0583	0.0478	0.0471	<b>0.0510</b>	0.0474
$n = 100$	0.0884	0.0898	0.0341	0.0390	0.0594	0.0615	0.0484	0.0525	<b>0.0494</b>	0.0552	0.0514
$n = 200$	0.0949	0.0944	0.0417	0.0590	0.0650	0.0492	0.0555	0.0494	<b>0.0495</b>	0.0552	0.0514
$n = 400$	0.0938	0.0382	0.0382	0.0611	0.0471	0.0537	0.0515	0.0485	0.0519	0.0542	<b>0.0494</b>
$n = 800$	0.0915	0.0354	0.0613	0.0437	0.0515	<b>0.0496</b>	0.0533	0.0522	0.0495	0.0505	0.0527
$n = 1600$	0.0886	0.0424	0.0595	0.0470	0.0527	0.0476	0.0511	<b>0.0502</b>	0.0489	0.0515	0.0507

**Table 2.** Entries represent empirical rejection probabilities of the prepivot ADF test with  $q = n^a$  for various values of  $a$ . The data generating process was ARMA(1,1) with  $(\phi, \theta) = (1, -0.5)$ .

$a =$	0.09	0.13	0.17	0.21	0.25	0.29	0.33	0.37	0.41	0.45	0.49
$n = 50$	0.1712	0.1675	0.1706	0.1001	0.0958	0.0708	0.0743	0.0576	0.0619	0.0541	<b>0.0489</b>
$n = 100$	0.1837	0.1844	0.1075	0.1027	0.0710	0.0767	0.0648	0.0619	<b>0.0501</b>	<b>0.0501</b>	0.0514
$n = 200$	0.1915	0.1953	0.1074	0.0778	0.0776	0.0635	0.0569	0.0522	<b>0.0503</b>	0.0528	0.0542
$n = 400$	0.1992	0.1083	0.1077	0.0748	0.0622	0.0571	0.0518	0.0520	<b>0.0495</b>	0.0527	0.0513
$n = 800$	0.2037	0.1128	0.0737	0.0680	0.0555	0.0553	<b>0.0503</b>	0.0490	0.0530	0.0534	<b>0.0502</b>
$n = 1600$	0.2118	0.1140	0.0797	0.0660	0.0568	0.0571	0.0517	<b>0.0515</b>	0.0550	0.0561	0.0472

**Table 3.** Entries represent empirical rejection probabilities of the prepivot ADF test with  $q = n^a$  for various values of  $a$ . The data generating process was ARMA(1,1) with  $(\phi, \theta) = (0.985, 0.5)$ .

$a =$	0.09	0.13	0.17	0.21	0.25	0.29	0.33	0.37	0.41	0.45	0.49
$n = 50$	<b>0.1150</b>	<b>0.1150</b>	0.1130	0.0851	0.0843	0.0958	0.0894	0.0902	0.0875	0.0890	0.0805
$n = 100$	<b>0.1752</b>	0.1741	0.1147	0.1078	0.1408	0.1288	0.1145	0.1229	0.1161	0.1172	0.1170
$n = 200$	<b>0.3044</b>	0.2986	0.1781	0.2271	0.2283	0.1949	0.2035	0.2037	0.1983	0.1887	0.1881
$n = 400$	<b>0.6175</b>	0.3756	0.3579	0.4897	0.4308	0.4494	0.4239	0.4117	0.4084	0.3781	0.3692
$n = 800$	<b>0.9593</b>	0.8139	0.8981	0.8501	0.8673	0.8525	0.8380	0.8265	0.8074	0.7712	0.7408
$n = 1600$	<b>1.0000</b>	0.9992	<b>1.0000</b>	0.9993	0.9987	0.9993	0.9986	0.9975	0.9961	0.9932	0.9855

**Table 4.** Entries represent empirical rejection probabilities of the prepivot ADF test with  $q = n^a$  for various values of  $a$ . The data generating process was ARMA(1,1) with  $(\phi, \theta) = (0.985, -0.5)$ .

$a =$	0.09	0.13	0.17	0.21	0.25	0.29	0.33	0.37	0.41	0.45	0.49
$n = 50$	0.1884	<b>0.1901</b>	0.1821	0.1243	0.1238	0.1000	0.1038	0.0942	0.0961	0.0939	0.0911
$n = 100$	0.3074	<b>0.3091</b>	0.1948	0.1973	0.1489	0.1546	0.1367	0.1226	0.1199	0.1241	0.1195
$n = 200$	<b>0.5455</b>	0.5417	0.3550	0.2777	0.2686	0.2420	0.2193	0.2076	0.1983	0.2001	0.1947
$n = 400$	<b>0.8786</b>	0.6832	0.6819	0.5686	0.4975	0.4629	0.4379	0.4171	0.4168	0.3876	0.3851
$n = 800$	<b>0.9989</b>	0.9813	0.9458	0.9182	0.8920	0.8736	0.8533	0.8419	0.8162	0.7875	0.7546
$n = 1600$	<b>1.0000</b>	<b>1.0000</b>	0.9999	0.9998	<b>1.0000</b>	0.9990	0.9984	0.9975	0.9974	0.9939	0.9861

**Table 5.** Entries represent empirical rejection probabilities of the prepivot ADF test with  $q = n^a$  for various values of  $a$ . The data generating process was ARMA(1,1) with  $(\phi, \theta) = (0.97, 0.5)$ .

$a =$	0.09	0.13	0.17	0.21	0.25	0.29	0.33	0.37	0.41	0.45	0.49
$n = 50$	0.1668	<b>0.1740</b>	0.1728	0.1069	0.1047	0.1253	0.1267	0.1115	0.1091	0.1085	0.1171
$n = 100$	<b>0.3033</b>	<b>0.3033</b>	0.1644	0.1700	0.2153	0.2139	0.1877	0.1936	0.1808	0.1898	0.1764
$n = 200$	<b>0.5977</b>	0.5960	0.3511	0.4549	0.4522	0.3890	0.4117	0.3889	0.3813	0.3607	0.3346
$n = 400$	<b>0.9600</b>	0.7989	0.7913	0.8787	0.8264	0.8441	0.8129	0.7859	0.7601	0.7357	0.6897
$n = 800$	<b>1.0000</b>	0.9978	0.9998	0.9985	0.9987	0.9976	0.9982	0.9950	0.9906	0.9832	0.9730
$n = 1600$	<b>1.0000</b>	0.9999									

**Table 6.** Entries represent empirical rejection probabilities of the prepivot ADF test with  $q = n^a$  for various values of  $a$ . The data generating process was ARMA(1,1) with  $(\phi, \theta) = (0.97, -0.5)$ .

$a =$	0.09	0.13	0.17	0.21	0.25	0.29	0.33	0.37	0.41	0.45	0.49
$n = 50$	0.2861	<b>0.2871</b>	0.2817	0.1854	0.1793	0.1536	0.1500	0.1308	0.1311	0.1225	0.1151
$n = 100$	<b>0.5246</b>	0.5194	0.3356	0.3326	0.2614	0.2654	0.2336	0.2076	0.2009	0.1856	0.1817
$n = 200$	0.8698	<b>0.8712</b>	0.6678	0.5514	0.5507	0.4752	0.4482	0.4060	0.3993	0.3376	0.3516
$n = 400$	<b>0.9987</b>	0.9762	0.9774	0.9421	0.9003	0.8723	0.8372	0.8102	0.7898	0.7592	0.7092
$n = 800$	<b>1.0000</b>	<b>1.0000</b>	0.9999	<b>1.0000</b>	0.9993	0.9994	0.9983	0.9975	0.9951	0.9870	0.9775
$n = 1600$	<b>1.0000</b>	0.9999	0.9998								

## 5. Bootstrap Selection of Tuning Parameters for Hypothesis Tests

In this section, we propose a state of the art tuning parameter selection algorithm for general hypothesis tests, and apply it to the case of lag length selection for the prepivot ADF test. The general algorithm proceeds as follows. Given time series data  $(X_1, \dots, X_n)$  from an assumed model, and a test statistic  $t_{n,q}$  where a tuning parameter  $q$  is involved, the procedure is as follows:

### Bootstrap Algorithm for Tuning Parameter Selection for Hypothesis Tests

1. Use an information-based choice of  $q$  from the sample  $(X_1, \dots, X_n)$  to obtain initial parameter estimates for the assumed model.
2. Use the estimated parameters to compute residuals that are not restricted to be under the null.
3. Bootstrap the residuals to construct  $B$  stretches of bootstrap data *under the null*.

4. Compute the test statistic  $t_{n,q}^*$  from each of the  $B$  bootstrap samples across the range of acceptable values of the tuning parameter. Collect the acceptance/rejections of these bootstrap tests as a matrix of 0s and 1s.
5. Select the tuning parameter that has empirical rejection probability closest to the nominal level of the test; this is essentially done by computing the column averages of the matrix generated in the previous step and selecting the index with column average closest to the nominal level. If there is more than one such choice, select the tuning parameter associated with the smallest fitted model.
6. The above is the optimal tuning parameter called  $q^*$ .
7. The optimal tuning parameter is fed back to the original data and the hypothesis test is performed with  $q^*$ , i.e., computing  $t_{n,q^*}$  on  $(X_1, \dots, X_n)$ .
8. Reject the null hypothesis if the  $p$  value of the test based on  $t_{n,q^*}$  is less than the nominal level.

Occam's razor postulates that if two models are equally apt in explaining a phenomenon at hand, then the smallest, i.e., least complex, model is preferable. Step 5 of the above algorithm encapsulates our preference for smaller, more parsimonious models. There are several instances in the literature of applying some form of bootstrap with the purpose of choosing a tuning parameter, see Leger and Romano [17], Fenga and Politis [18] and Shao [19]. However, it is important to note that our bootstrap algorithm of Section 5 is novel. As far as we know, no other work has proposed to generate bootstrap samples that obey the null (whether the data obey the null or not) with the purpose of choosing the smallest model order that still yields size close to nominal. In the next subsection, we will apply this algorithm to the ADF test in which case, as we have shown, using a smaller model generally leads to better power.

### 5.1. ADF with Bootstrap-Assisted Lag Choice

In this subsection, we apply the above general bootstrap algorithm for tuning parameter selection to the pre pivoted ADF test. We perform numerical experiments to empirically verify the size and power properties of our algorithm and report the results. As per the work of Paparoditis and Politis [6], we observe that the Akaike Information Criterion (AIC) works reasonably well with regards to both size and power of the ADF test in the case of positive MA parameters. In the case of negative MA parameters, the AIC based tests suffer from over-rejection of the null. The Modified AIC (MAIC) due to Ng and Perron [7] was designed specifically to ameliorate the ADF test's over-rejection of the null in the presence of negative MA parameters. Their method uses a combination of GLS detrending along with a modified information criterion to select the optimal lag length for the ADF test. The MAIC is designed to favor larger lag lengths in the presence of a negative MA parameter, but as shown by Paparoditis and Politis [6], this can hurt the power performance of the test due to the asymptotic collinearity effect.

Our lag length selection algorithm, dubbed the 'bootstrap-assisted lag choice' (BALC) is given below; it uses the residual AR bootstrap of Paparoditis and Politis [9] along with the pre pivoting idea of Beran [8] with the aim of selecting lag lengths that yield both good size and power. We performed numerical experiments using an ARMA(1,1) data generating process with six different combinations of the ARMA parameters  $\phi$  and  $\theta$ . For each ARMA(1,1) configuration and sample size  $n$ , we generated 10,000 different time series and recorded the performance of our novel lag length selection algorithm. This allowed us to compute the empirical rejection probabilities corresponding to our algorithm at the nominal level 0.05. We compare the results of our novel lag length selection method with pre pivoted ADF tests with lag lengths selected by AIC and MAIC optimality. The experimental design is as follows:

#### Bootstrap Assisted Lag Choice Algorithm for the ADF Test

1. We are given a stretch of time series data  $X_1, \dots, X_n$  to be tested for a possible unit root.
2. Run the ADF regression (1) on this data with lag length  $q$  selected by AIC. AIC minimization is done over the range 1 to  $\sqrt{n}$ . Denote the selected lag by  $q_{AIC}$ .

3. Use the estimated parameters and centered residuals from the ADF regression to construct  $B_1$  bootstrap samples under the null, using the AR bootstrap of Paparoditis and Politis [9].
4. For each of the bootstrap samples, perform a prepivot ADF test with varying lag lengths. The bootstrap methodology for the prepivot test also uses the AR bootstrap of Paparoditis and Politis [9], and generates  $B_2$  bootstrap samples for each  $q$ . The range of values for  $q$  we chose was from  $[0.5 \cdot q_{AIC}]$  to  $[1.5 \cdot q_{AIC}]$ . The result of these tests is stored in a 0–1 matrix with  $B_1$  rows and appropriate number of columns.
5. The optimal  $q^*$  is then picked by using the value of  $q$  that yields smallest type-1 error for all of the bootstrap samples i.e., the value of  $q$  that yields column average closest to the nominal level 0.05.
6. We then perform a prepivot ADF test with  $B_2$  bootstrap samples on the original data with number of lags equal to  $q^*$ .
7. The unit root null hypothesis is rejected if the  $p$ -value is less than the nominal level 0.05.

### 5.2. Discussion of Simulation Results

The Tables below list the empirical rejection probabilities of the prepivot ADF test with lag lengths selected by three different data dependent approaches. We list the results of our algorithm and compare them with those with lags selected by AIC, BIC and MAIC optimality. AIC optimization is carried out over lag lengths  $q$  ranging from 1 to  $\sqrt{n}$ , while MAIC optimization is carried out over  $q$  ranging from 1 to  $[12 \cdot (\frac{n+1}{100})^{0.25}]$  as per Schwert [3]. For our experiments, the values of  $B_1, B_2$  were set to 250 and 100 respectively in the interest of computational parsimony since our simulation involved 10,000 replications; in practice, given a single dataset in hand, the values of  $B_1, B_2$  can easily become of the order of 1000. In terms of computational efficiency, for a single dataset, the computational complexity is entirely determined by the values of  $B_1$  and  $B_2$  and the lag length initially chosen by AIC. More specifically, the ADF regression is performed is  $1 + B_1 B_2 q_{AIC} + B_2$  times. Step 4 from the algorithm is highly conducive to parallelization—the search can be run over all candidate values for  $q^*$  at the same time since the results from one candidate value of  $q^*$  are independent of those for other candidate values—thereby improving the computational efficiency and runtime of the algorithm.

Tables 7 and 8 correspond to the setting wherein the data  $\{X_t\}$  have a unit root. The entries therefore list the rejection probabilities under the null, and should ideally be as close to the nominal level 0.05 as possible. For each sample size in each Table, the rejection probability closest to the nominal level is presented in bold.

Putting together the results of Tables 7 and 8, we see that the performance of our method is comparable to that of AIC and MAIC. It is clear from the Tables that tests achieve more accurate size for larger sample sizes. Not surprisingly, the bootstrap selection method also suffers from over-rejection of the null hypothesis in the presence of negative MA parameter.

Tables 9–12 correspond to the setting wherein the data  $\{X_t\}$  obey the alternative hypothesis. The entries therefore correspond to the power of the prepivot ADF test, and should be as high as possible. For each sample size in each Table, the highest rejection probability is presented in bold.

These results allow us to draw the following conclusions. First, it appears that our method performs reasonably well with respect to finite sample power. The power of the BALC test is higher than the power of the prepivot test with AIC or MAIC lag lengths. We also note that the method is consistent as its power approaches 1 with increasing sample size. Additionally, we observe that Tables 10–12 show significant increases in the power of our method compared to the two information based criteria. What is interesting is that the biggest improvements are seen for intermediate sample sizes. In particular, we observe that our method outperforms the MAIC vis-à-vis power. This can primarily be attributed to the fact that our method is designed to favor shorter lags (provided their associated

size is close to nominal), thereby avoiding the asymptotic collinearity problem under the alternative hypothesis. It appears that our proposed bootstrap-assisted lag choice (BALC) generally outperforms AIC and MAIC in the cases of size and power, with the exception that the MAIC achieves better size in the well-known problematic case of a negative MA parameter. The prepivoted test with lag length selected by BIC provides a comparable but slightly higher power in the presence of a negative MA parameter as compared to our BALC based test.

All in all, it appears that the BALC method works well at arriving at a compromise between size and power performance. Our lag selection method yields accurate size as well as improved power over the information based lag length selection criteria. The lag length selection idea along with prepivoting appears to yield appreciable improvements over the asymptotic ADF test.

**Table 7.**  $(\phi, \theta) = (1, 0.5)$ .

$n =$	50	100	200	400	800	1600
AIC	0.0578	0.0512	<b>0.0511</b>	0.0522	0.0526	0.0544
BIC	0.0706	0.0644	0.0537	<b>0.0499</b>	0.0543	0.0526
MAIC	<b>0.0465</b>	<b>0.0500</b>	0.0463	0.0497	0.0541	0.0506
BALC	0.0612	0.0582	0.0541	0.0526	<b>0.0514</b>	<b>0.0495</b>

**Table 8.**  $(\phi, \theta) = (1, -0.5)$ .

$n =$	50	100	200	400	800	1600
AIC	0.0964	0.0864	0.0714	0.0656	0.0593	<b>0.0528</b>
BIC	0.1152	0.1286	0.1101	0.0945	0.0782	0.0704
MAIC	<b>0.0673</b>	<b>0.0650</b>	<b>0.0631</b>	<b>0.0602</b>	<b>0.0588</b>	<b>0.0528</b>
BALC	0.1174	0.1056	0.1049	0.0966	0.0842	0.0761

**Table 9.**  $(\phi, \theta) = (0.985, 0.5)$ .

$n =$	50	100	200	400	800	1600
AIC	0.0931	0.1166	0.1892	0.4162	0.8375	0.9987
BIC	<b>0.1008</b>	<b>0.1232</b>	0.1928	0.3812	0.8402	<b>0.9995</b>
MAIC	0.0893	0.1133	0.1965	0.4239	<b>0.8462</b>	0.9983
BALC	0.0965	0.1215	<b>0.2113</b>	<b>0.4391</b>	0.8407	0.9989

**Table 10.**  $(\phi, \theta) = (0.985, -0.5)$ .

$n =$	50	100	200	400	800	1600
AIC	0.1199	0.1586	0.2429	0.4837	0.8712	0.9991
BIC	<b>0.1432</b>	<b>0.2207</b>	<b>0.3348</b>	<b>0.5862</b>	<b>0.9290</b>	<b>0.9998</b>
MAIC	0.1067	0.1395	0.2313	0.4707	0.8678	0.9985
BALC	0.1378	0.2049	0.3245	0.5769	0.8972	0.9992

**Table 11.**  $(\phi, \theta) = (0.97, 0.5)$ .

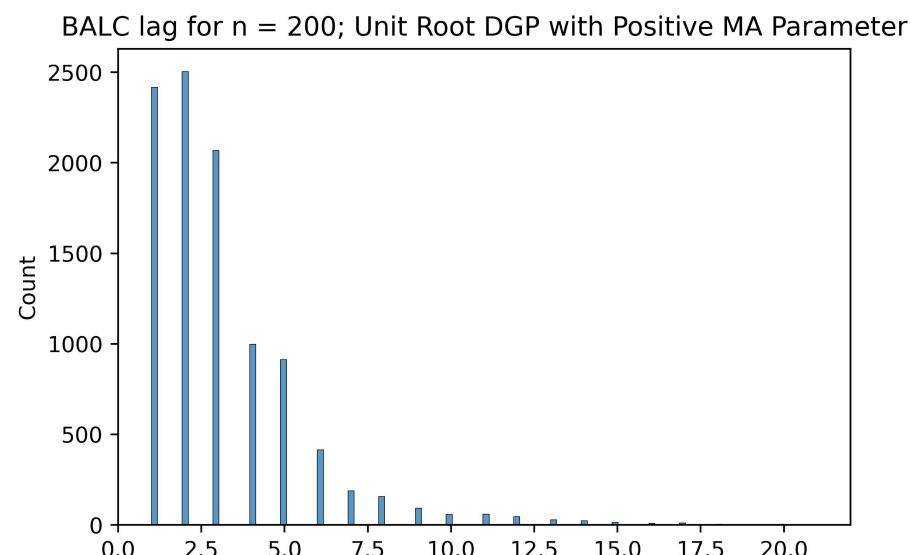
$n =$	50	100	200	400	800	1600
AIC	0.1142	0.1829	0.3786	0.7974	0.9955	<b>1.0000</b>
BIC	0.1286	0.2100	0.3902	0.7965	0.9958	<b>1.0000</b>
MAIC	0.1127	0.1819	0.3799	0.7910	0.9947	0.9999
BALC	<b>0.1364</b>	<b>0.2198</b>	<b>0.4336</b>	<b>0.8205</b>	<b>0.9963</b>	<b>1.0000</b>

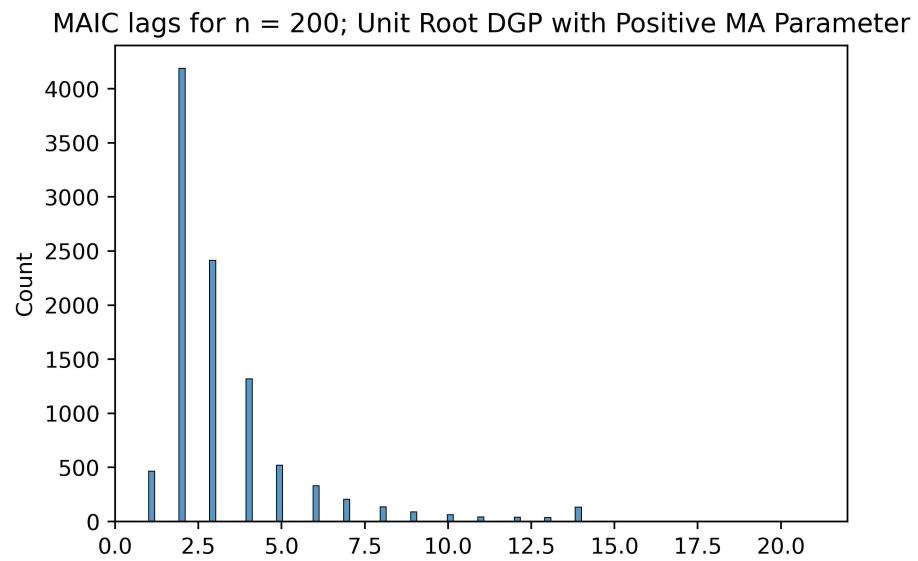
**Table 12.**  $(\phi, \theta) = (0.97, -0.5)$ .

$n =$	50	100	200	400	800	1600
AIC	0.1689	0.2713	0.4869	0.8521	0.9980	<b>1.0000</b>
BIC	<b>0.2218</b>	<b>0.3721</b>	<b>0.6242</b>	<b>0.9299</b>	<b>0.9997</b>	<b>1.0000</b>
MAIC	0.1530	0.2242	0.4499	0.8343	0.9940	<b>1.0000</b>
BALC	0.2071	0.3409	0.5874	0.8959	0.9993	<b>1.0000</b>

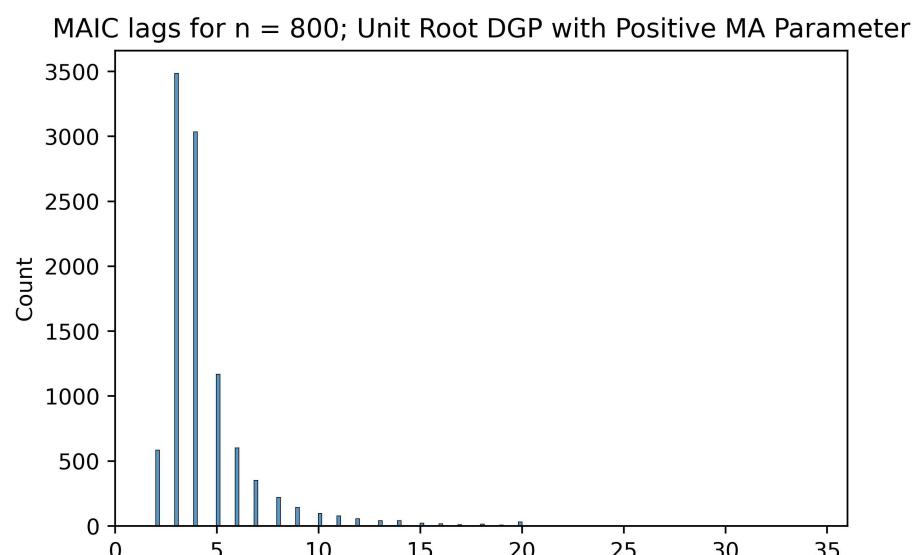
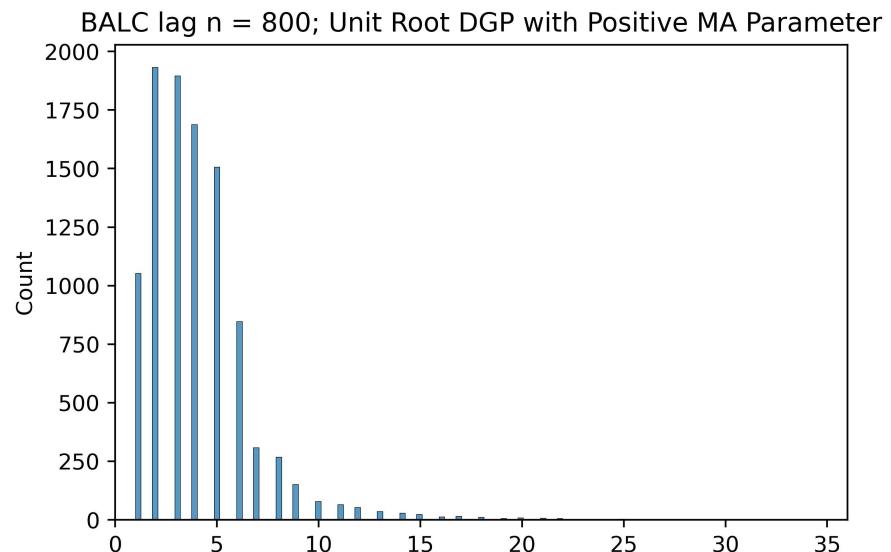
### 5.3. Comparing the BALC Lag Lengths to MAIC

Given below are histograms comparing the optimal lags selected by our BALC algorithm and lags selected by MAIC optimality. They serve as evidence of the fact that the BALC tends to pick lower lags than MAIC, while still searching in a range higher than the lags selected by AIC—therefore splitting the difference between AIC and MAIC. The lag lengths selected by BALC are generally shorter than those selected by MAIC. Although there is a stochastic component to the BALC framework, we observe that the choice to cut off the search for lags at  $1.5 * q_{\text{AIC}}$  allows us to inflate the search space from just the value selected by AIC while simultaneously preventing the asymptotic collinearity problem. The asymptotic collinearity problem stems from the fact that as the number of lags  $q$  increases, the regressors in the ADF regression become asymptotically collinear—this is the reason for the loss in power of the ADF test as pointed out in Paparoditis and Politis (2018) [6]. If a regression is collinear, then the OLS estimator is ill-conditioned/non-unique; fortunately, here we have only approximate/asymptotic collinearity (under the alternative) whose effect can be mitigated by choosing the smallest  $q$  possible. This issue is demonstrated in Figures 1–4 and best visualized in Figures 5 and 6, where the MAIC picks fairly large lag lengths for a significant proportion of the 10,000 DGPs whereas the BALC algorithm avoids larger lag lengths.

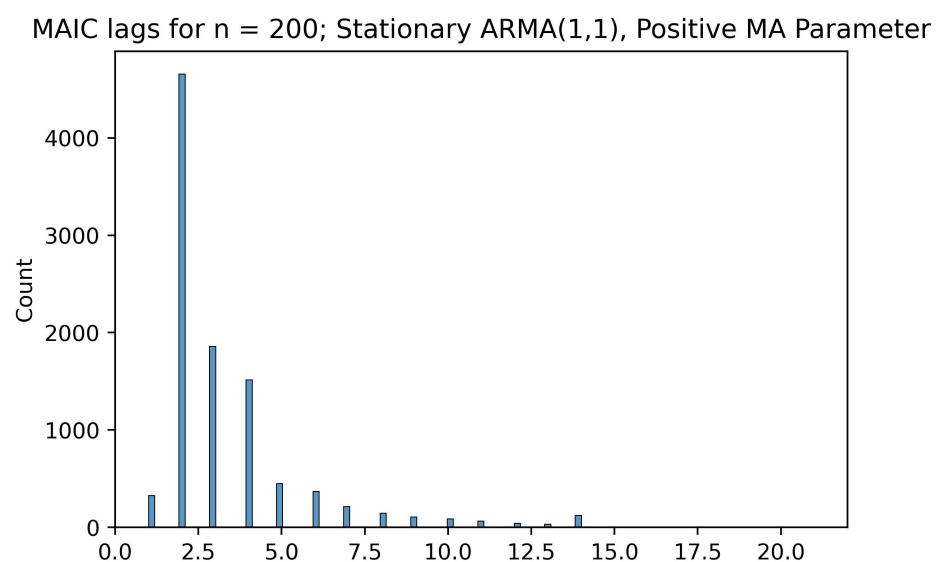
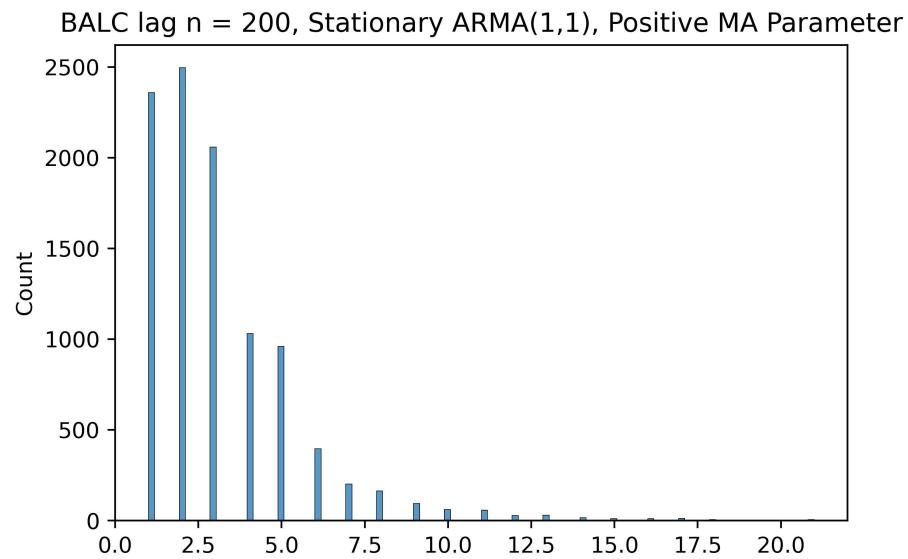
**Figure 1.** *Cont.*



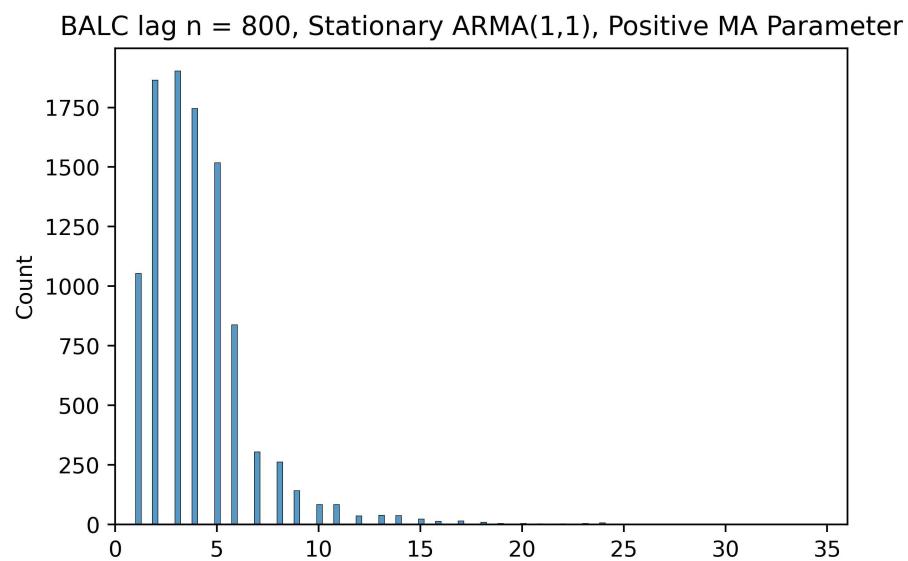
**Figure 1.** Comparison of lag lengths selected by BALC and MAIC for  $(\rho, \theta) = (1, 0.5)$ ,  $n = 200$ .



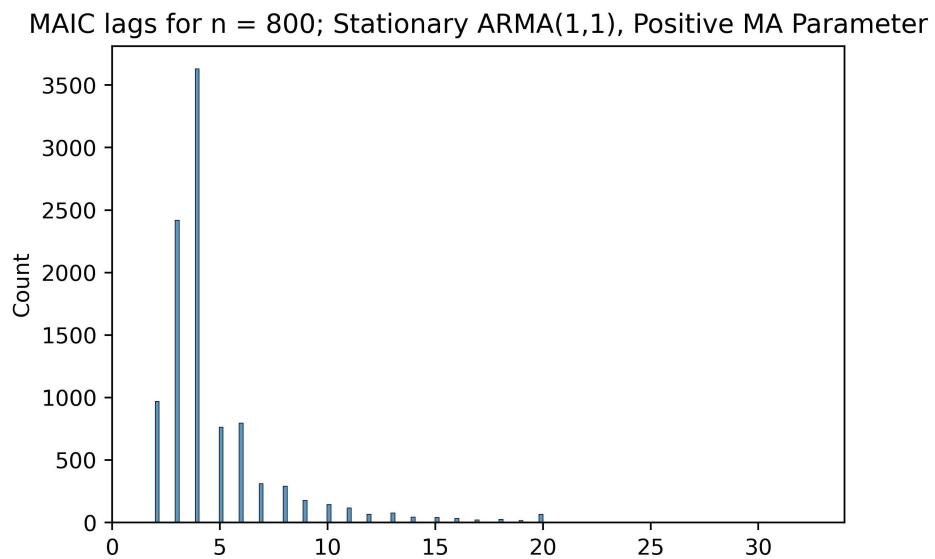
**Figure 2.** Comparison of lag lengths selected by BALC and MAIC for  $(\rho, \theta) = (1, 0.5)$ ,  $n = 800$ .



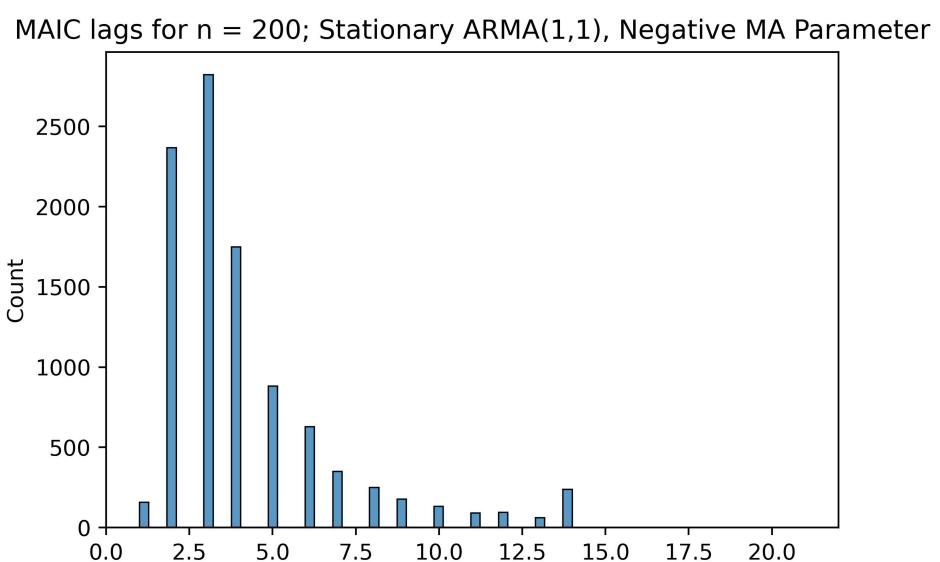
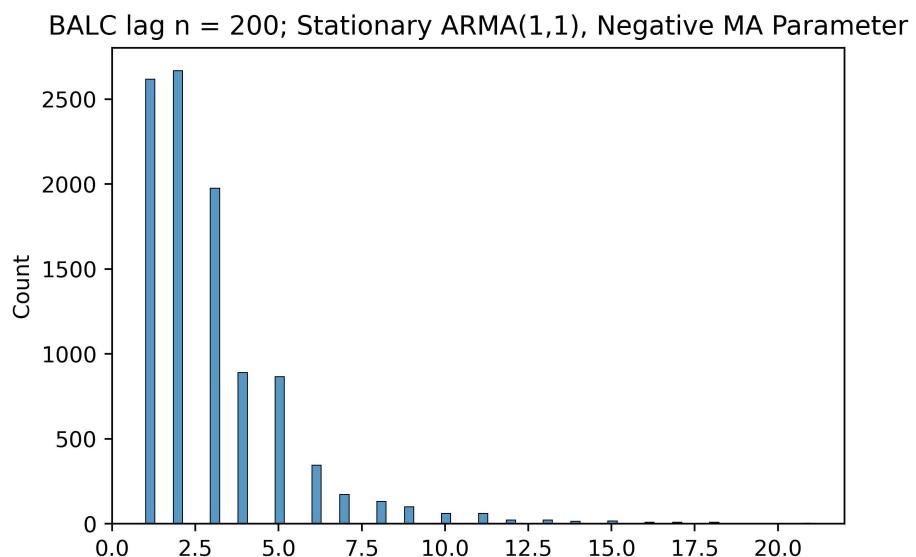
**Figure 3.** Comparison of lag lengths selected by BALC and MAIC for  $(\rho, \theta) = (0.985, 0.5)$ ,  $n = 200$ .



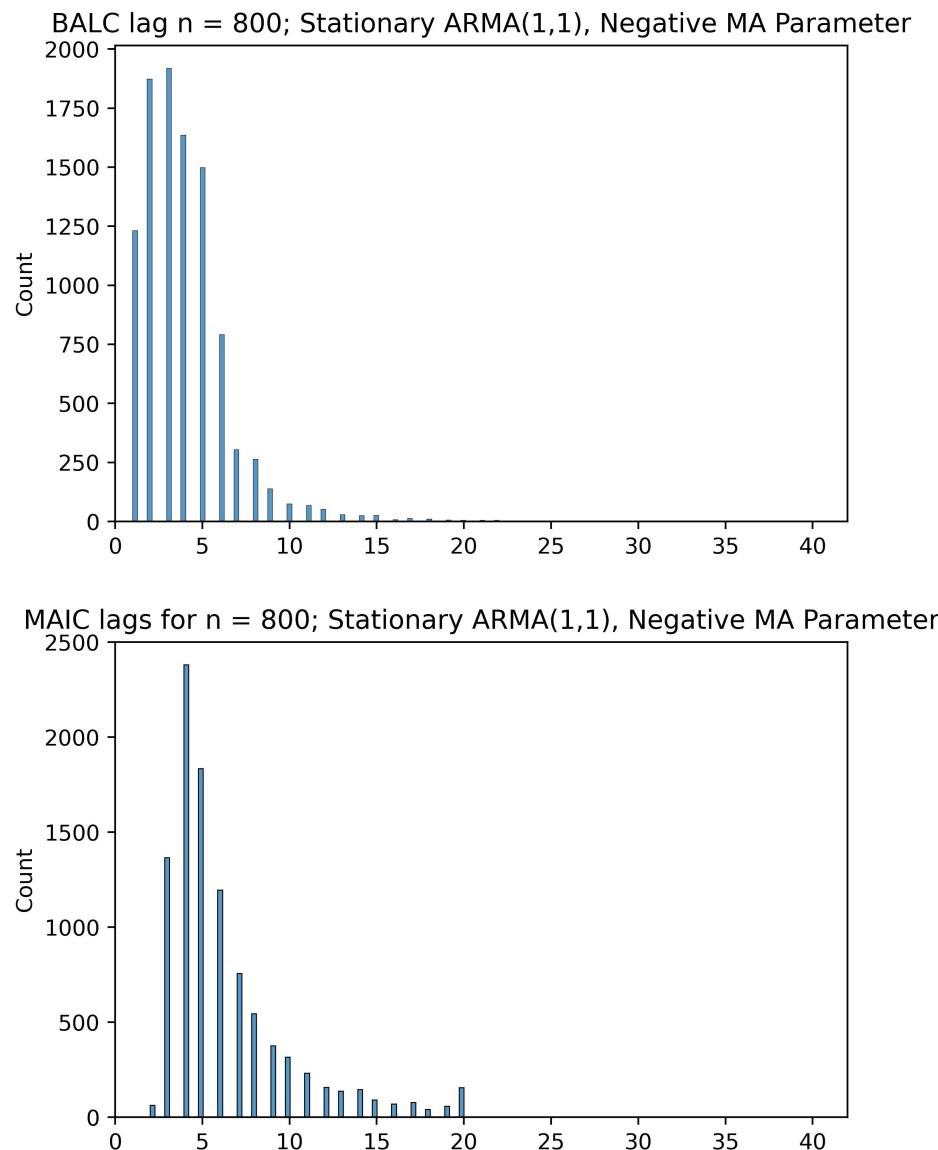
**Figure 4. Cont.**



**Figure 4.** Comparison of lag lengths selected by BALC and MAIC for  $(\rho, \theta) = (0.985, 0.5)$ ,  $n = 800$ .



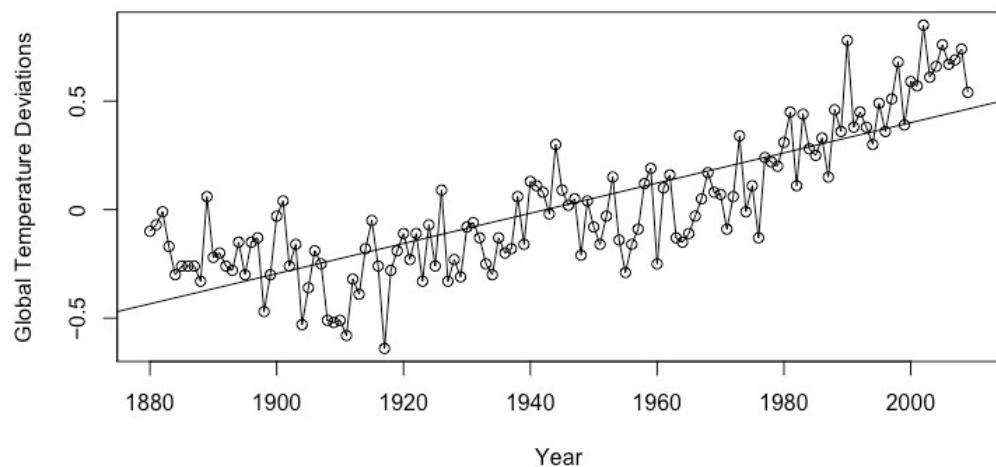
**Figure 5.** Comparison of lag lengths selected by BALC and MAIC for  $(\rho, \theta) = (0.97, -0.5)$ ,  $n = 200$ .



**Figure 6.** Comparison of lag lengths selected by BALC and MAIC for  $(\rho, \theta) = (0.97, -0.5)$ ,  $n = 800$ .

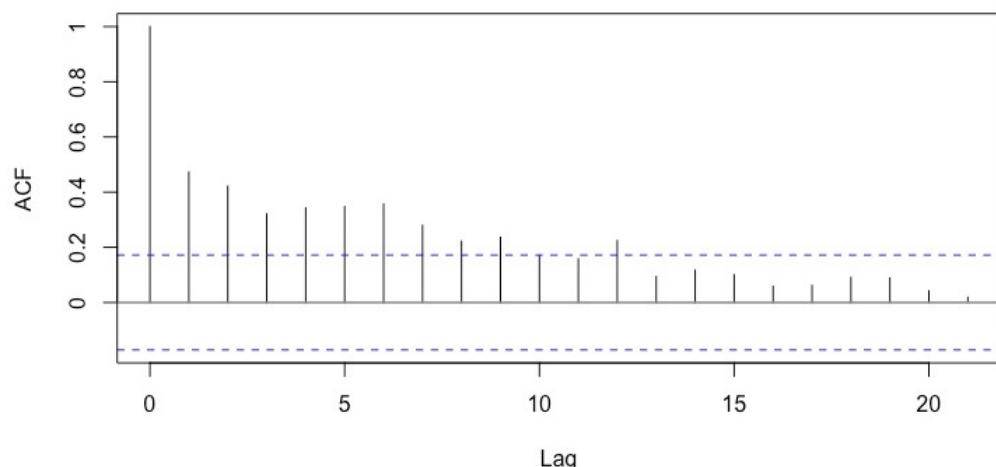
#### 5.4. A Real Data Application

Lastly, we apply our novel lag selection method to a real data example. We apply our BALC testing procedure to the dataset of Figure 7 that is discussed in the textbook on time series analysis by Shumway and Stoffer (2017) [20]. The dataset represents the yearly average global temperature deviations from 1880 to 2009, with the deviations measured in degrees Celsius with respect to the average temperature between 1951 and 1980. In Section 3 of Paparoditis and Politis [6], the authors comment on the failure of the `tseries` implementation of the ADF test to reject the null, as a consequence of selecting too many lags. As per their discussion and diagnostic figures, there is strong evidence that the *detrended* data do not contain any strong evidence for a unit root, and therefore obeys the alternative hypothesis—evidenced by the ACF plot of the detrended data given in Figure 8. The BALC method yields an optimal lag length of 3, compared to the `tseries` ADF function's recommendation of 5. The prepivoted ADF test with lag length  $q = 3$  then rejects the unit root null hypothesis, which is the expected result. In contrast, the asymptotic ADF test with  $q = q_{AIC}$  (which also equals 5) fails to reject the null.



**Figure 7.** Yearly average global temperature deviations data with superimposed fitted linear trend.  $n = 130$ .

#### ACF of Detrended Data



**Figure 8.** Correlogram of detrended dataset.

#### 6. Discussion

In this paper, we have presented a novel bootstrap assisted lag length selection criterion for the ADF test. As the asymptotic theory shows, the choice of lag length for the ADF test significantly affects its size and power properties. The convergence rate of the ADF estimator is different under the null and under the alternative. In addition, there is a tug-of-war: we need a large  $q$  to ensure size close to nominal while we need a smaller  $q$  (still diverging though) to have good power. The practical implementation of the ADF test requires the practitioner to choose the  $q$  to use in the ADF test without knowing whether the null or the alternative holds as this is the very question of the test. Given this difficult situation, our paper proposes a bootstrap-based algorithm that helps the practitioner choose a value of  $q$  that is the smallest among all  $q$  values that achieve size close to nominal. Our Tables 7–12 show that our algorithm works well in practice as it automatically adapts to the underlying data structure without the need for the practitioner to choose  $q$  in an ad hoc way, e.g., by minimizing a criterion such as AIC, BIC, MAIC, etc.

**Author Contributions:** Conceptualization, S.M. and D.N.P.; Methodology, S.M. and D.N.P.; Software, S.M. and D.N.P.; Formal analysis, S.M. and D.N.P.; Investigation, S.M. and D.N.P.; Writing—original draft, S.M. and D.N.P.; Visualization, S.M. and D.N.P.; Supervision, D.N.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research of the first author was partially supported by the Richard Libby Graduate Research Award. The research of the second author was partially supported by NSF Grant DMS 24-13718.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

**Acknowledgments:** This research was done using services provided by the OSG Consortium [21–24] which is supported by the National Science Foundation awards #2030508 and #1836650.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Dickey, D.A.; Fuller, W.A. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.* **1979**, *74*, 427–431.
2. Phillips, P.C.; Perron, P. Testing for a unit root in time series regression. *Biometrika* **1988**, *75*, 335–346. [\[CrossRef\]](#)
3. Schwert, G.W. Tests for unit roots: A Monte Carlo investigation. *J. Bus. Econ. Stat.* **2002**, *20*, 5–17. [\[CrossRef\]](#)
4. Perron, P.; Ng, S. Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. *Rev. Econ. Stud.* **1996**, *63*, 435–463. [\[CrossRef\]](#)
5. Davidson, R.; MacKinnon, J.G. *Econometric Theory and Methods*; Oxford University Press: New York, NY, USA, 2004; Volume 5.
6. Paparoditis, E.; Politis, D.N. The asymptotic size and power of the augmented Dickey–Fuller test for a unit root. *Econom. Rev.* **2018**, *37*, 955–973. [\[CrossRef\]](#)
7. Ng, S.; Perron, P. Lag length selection and the construction of unit root tests with good size and power. *Econometrica* **2001**, *69*, 1519–1554. [\[CrossRef\]](#)
8. Beran, R. Preprinting test statistics: A bootstrap view of asymptotic refinements. *J. Am. Stat. Assoc.* **1988**, *83*, 687–697. [\[CrossRef\]](#)
9. Paparoditis, E.; Politis, D.N. Bootstrapping unit root tests for autoregressive time series. *J. Am. Stat. Assoc.* **2005**, *100*, 545–553. [\[CrossRef\]](#)
10. Said, S.E.; Dickey, D.A. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* **1984**, *71*, 599–607. [\[CrossRef\]](#)
11. Chang, Y.; Park, J.Y. On the asymptotics of ADF tests for unit roots. *Econom. Rev.* **2002**, *21*, 431–447. [\[CrossRef\]](#)
12. Brockwell, P.J.; Davis, R.A. *Time Series: Theory and Methods*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1991.
13. Kreiss, J.P.; Paparoditis, E.; Politis, D.N. On the range of validity of the autoregressive sieve bootstrap. *Ann. Stat.* **2011**, *39*, 2103–2130. [\[CrossRef\]](#)
14. Hamilton, J.D. *Time Series Analysis*; Princeton University Press: Princeton, NJ, USA, 2020.
15. Politis, D.N.; Romano, J.P.; Wolf, M. *Subsampling*; Springer: Berlin/Heidelberg, Germany, 1999.
16. Aylar, E.; Smeekes, S.; Westerlund, J. Lag truncation and the local asymptotic distribution of the ADF test for a unit root. *Stat. Pap.* **2019**, *60*, 2109–2118. [\[CrossRef\]](#)
17. Léger, C.; Romano, J.P. Bootstrap choice of tuning parameters. *Ann. Inst. Stat. Math.* **1990**, *42*, 709–735. [\[CrossRef\]](#)
18. Fenga, L.; Politis, D.N. LASSO order selection for sparse autoregression: A bootstrap approach. *J. Stat. Comput. Simul.* **2017**, *87*, 2668–2688. [\[CrossRef\]](#)
19. Shao, J. Bootstrap model selection. *J. Am. Stat. Assoc.* **1996**, *91*, 655–665. [\[CrossRef\]](#)
20. Shumway, R.H.; Stoffer, D.S. *Time Series Analysis and Its Applications: With R Examples*; Springer: Berlin/Heidelberg, Germany, 2017.
21. Pordes, R.; Petravick, D.; Kramer, B.; Olson, D.; Livny, M.; Roy, A.; Avery, P.; Blackburn, K.; Wenaus, T.; Würthwein, F.; et al. The open science grid. *J. Phys. Conf. Ser.* **2007**, *78*, 012057. [\[CrossRef\]](#)
22. Sfiligoi, I.; Bradley, D.C.; Holzman, B.; Mhashilkar, P.; Padhi, S.; Wurthwein, F. The pilot way to grid resources using glideinWMS. In Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering, Los Angeles, CA, USA, 31 March–2 April 2009; Volume 2, pp. 428–432. [\[CrossRef\]](#)
23. OSG. *OSPool*; OSG, 2006. Available online: [https://osg-htc.org/services/open\\_science\\_pool.html](https://osg-htc.org/services/open_science_pool.html) (accessed on 8 October 2024). [\[CrossRef\]](#)
24. OSG. *Open Science Data Federation*; OSG, 2015. Available online: <https://osdf.osg-htc.org/> (accessed on 8 October 2024). [\[CrossRef\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.