行政院國家科學委員會補助專題研究計畫成果報告

改善自我迴歸模型之統計推論:簡單函數轉換的應用

計畫類別： 個別型計畫

計畫編號：NSC 89-2415-H-004-036-

執行期間：89 年 8 月 1 日至 91 年 10 月 31 日

計畫主持人：郭炳伸

本成果報告包括以下應繳交之附件：

□赴國外出差或研習心得報告一份

□赴大陸地區出差或研習心得報告一份

□出席國際學術會議心得報告及發表之論文各一份

□國際合作研究計畫國外研究報告書一份

執行單位：政治大學國貿系

中 華 民 國 91 年 10 月 30 日

Bootstrap Inference for Unit Root with Dependent Errors *

Project NSC 89-2415-H-004-036-
1/8/1999 - 31/7/2002
Biing-Shen Kuo †
Department of International Trade, National Chengchi University

Abstract

This paper proposes the bootstrap procedures for testing for the presence of a unit root in an AR model with weakly dependent errors. Our bootstrap proposal is based on an autoregressive approximation of order increasing with sample size to the model. We establish the validity of such bootstrap approximations for the the limiting distribution of the test statistics considered. These statistics include commonly-used ADF, $DF-GLS$, MZ and $MZ-GLS$. We show as well that estimates of the bootstrap long-run variance or regression coefficients when the model allows for intercept and time trend are equivalent to their asymptotic counterparts. At little cost of power loss, our resampling scheme yields satisfactory control over the rejection probability for most of very small sample sizes in simulations.

JEL classification: C12, C14, C15, C22
Keywords: unit root test, bootstrap, AR model

*The project was originally entitled "Pivoting to Improve Accuracy of Confidence Intervals for Autoregressive Coefficient Near Unit Root". However, the ideas along the line as those suggested in the project application was later found to be infeasible and hard to be pursued. The project, then, was re-geared toward the direction relevant to what is being reported here. The change in the title of the project thus is simply a reflection of what the report is really about.

†Correspondence: Tel: (886 2) 2939 3091 ext. 81029, Fax: (886 2) 2938 7699, Email: bsku@nccu.edu.tw

1 Introduction

Testing for the presence of a unit root appears to be becoming a common empirical practice in economic time series analysis. The need for such a testing strategy may partly arise from the implication of economic theories, such as the purchasing power parity hypothesis and the persistence theory of innovations to the GDP, to the data. Partly, this is because the presence of a unit root in the series can invalidate statistical inference without paying attention to the non-standard asymptotic nature of the unit root statistics. Care needs to be exercised accordingly if a unit root is detected. The last decade has thus seen a booming development of the unit root literature, either on econometric techniques or on their applications.

The ongoing research on the econometric methods has been more or less geared toward improving inference from the extant unit root test statistics. The improvement in the finite samples can be achieved by both reducing size distortion through a better selection of lag orders (Ng and Perron; 1995, 2001), robust estimates of nuisance parameters (Ng and Perron, 1998) or small-sample adjustments to the test statistics (Perron and Ng, 1996), and gaining power through an efficient removal of trends from the data (Elliot, Rothenberg and Stock, 1996; and Ng and Perron, 2001).

We in this paper make the same effort to improve inference on unit root by taking the bootstrapping approach. Like most of the earlier contributions, the bootstrap proposal we set out for the unit root testing is to yield better estimates of the finite-sample critical values. For that, the bootstrap however directly works on estimating the distribution functions of the test statistics, generating estimates of the finite-sample critical values as a result. This route seems to be natural to work on, to the extent that the probabilistic information is contained in the distribution function. We bootstrap the test statistics that have been subject to intensive studies, specifically the classes of DF test statistics by Dicky and Fuller (1975) and Said and Dicky (1984) and Z test statistics by Phillips and Perron (1988) and subsequently modified by Perron and Ng (1996). The focal point here is on the version of these statistics with GLS detrending (see Elliot *et al.* (1996) and Ng and Perron (2001). While the modified statistics attain satisfactory finite-sample performance over their previous versions, we demonstrate in the experimental simulations that, with their bootstrap counterparts, there still has room to further reduce size bias for sample sizes that often encounter in practice or even smaller, at the same time maintaining comparable power to those with size-adjusted critical values. This proves to be remarkable in the situation with the presence of negative moving-average innovations, known to pose difficulty to the unit root testing (see e.g. Perron and Ng, 1996).

To justify our resampling algorithm, we establish the consistency of the bootstrap distribution of the test statistics. The asymptotic validity however is not obtained at the cost of limiting the usefulness of the bootstrap procedures. Our schemes are applicable to a wide spectrum of applications since the model under study is built on the condition allowing for infinite-order moving average error processes, including finite order $ARMA$ as special case. We borrow the idea of autoregressive approximation from Bühlmann (1997) to reproduce samples. To deal with the dependence properly in the data, as Horowitz (2000) emphasizes, is crucial to have the bootstrap well-perform. Phillips (2001) particularly shows that the bootstrap that leaves aside the dependence would turn a spurious regression into a cointegrating regression. The present results could thus be seen as an important extension of the work by Nankervis and Savin (1996) and Ferretti and Romo (1996) that consider only *iid* or $AR(1)$ errors.

Furthermore, the presence of the deterministic trend in the model, while rendering more reality, complicates the asymptotic argument for the bootstrap distribution consistency. The argument for simple random-walk model can not be readily carried over to the model with trend we consider here. As a by-product, we show the asymptotic equivalence of both the bootstrap trend coefficient and the bootstrap spectral density estimate at frequency zero.

The plan of the paper is as follows. In Section 2, we spell out the unit root test statistics under study. Section 3 describes our bootstrap resampling proposal. The consistency is provided and discussed in Section 4 for the bootstrap test distributions and parameters. Section 5 investigates the small-sample performance of the bootstrap proposal using Monte-Carlo simulation. Both $AR(1)$ and $MA(1)$ errors are considered. The last section contains concluding remarks. Proofs of the theorems and lemmas are given in the appendix.

2 Test Statistics

The model under study is an autoregressive model with intercept and trend. The model can expressed either in an unobserved component form or in a reduced (regression) form. For the former, we consider a series $\{y_t\}_{t=0}^T$ that is generated by:

$$y_t = \sum_{i=0}^{m} \mu_i t^i + u_t + \eta_t, \quad u_t = \alpha u_{t-1} + v_t \tag{1}$$

$$v_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}, \quad \psi_0 = 1. \tag{2}$$

where $\epsilon_t \overset{iid}{\sim} (0, \sigma_\epsilon^2)$ with an unknown distribution F. Let $\Psi(L) = 1 + \sum_{i=1}^{\infty} \psi_i L^i$. We further assume the error process $\{v_t\}$ is invertible, i.e. $\Psi(L)$ is non-zero on unit circle, and $\sum_{i=0}^{\infty} i|\psi_i| < \infty$. The class of error processes considered therefore includes the stationary and invertible ARMA process as a special sub-class. Under these assumptions, it is known that $\{v_t\}$ can have an infinite order autoregressive representation: $v_t = \sum_{j=1}^{\infty} \phi_j v_{t-j} + \epsilon_t$, where $\Phi(L) = \Psi(L)^{-1} = 1 + \sum_{j=1}^{\infty} \phi_j L^j$. This autoregressive representation of the errors indeed motivates our resampling scheme proposed subsequently. The spectral density at frequency zero of v_t or the so-called long-run variance then is defined to be $\sigma^2 = \sigma_\epsilon^2 \Psi(1)$. Besides, we assume the weak convergence holds for the partial sum of errors, $T^{-1/2} \sum_{t=1}^{rT} v_t \Rightarrow \sigma W(r)$. The model in (1) allows for deterministic component when setting $m = 0$ or 1.

The reduced form of the model is given by

$$y_t = \sum_{i=0}^{m} \beta_i t^i + \alpha y_{t-1} + \sum_{j=1}^{\infty} \phi_j \Delta y_{t-j}$$

with

$$\beta_0 = \mu_0(1-\alpha) + \mu_1(\alpha - \sum_{j=1}^{\infty} \phi_j) \text{ and } \beta_1 = \mu_1(1-\alpha). \quad (3)$$

This regression indeed is the well-known augmented Dick-Fuller regression if the error in (2) instead is assumed to be an $AR(p)$ process.

The null hypothesis of interest is to testing for the presence of a unit root, i.e. testing $H_0 : \alpha = 1$ against $H_a : \alpha < 1$. Many unit root test statistics have been proposed in the literature. Among them, the most popular choices appear to the ADF test by Said and Dicky (1984), $DF - GLS$ test by Elliot, Rothenberg and Stock (1996), Z test by Phillips and Perron (1988) and MZ test by Perron and Ng (1996). The MZ test is a modified version of the Z test, and proves to have a better size performance in the small samples. Furthermore, to enhance power as in Elliot *et al.*, Ng and Perron (2001) apply the idea of the GLS detrending to the class of MZ test statistics. These test statistics will be under investigation for their bootstrap counterparts to improve inference on the unit root hypothesis. Particularly, we focus on the test statistics under the GLS detrending. [1]

2.1 ADF and $DF - GLS$ tests

The ADF test is the t-statistic for $a_0 = 0$ in the following autoregression

$$\Delta \tilde{y}_t = a_0 \tilde{y}_{t-1} + \sum_{i=1}^{p} a_i \Delta \tilde{y}_{t-i} + e_{tp} \quad (4)$$

where the OLS detrended series $\tilde{y}_t = y_t - \sum_{i=0}^{m} \tilde{\beta}_i t^i$, in which $\tilde{\beta}_i$ is the OLS estimates for the deterministic trend coefficients. To implement the DF test, it requires a choice of the autoregressive truncation lag, p that is in practice determined by using AIC or BIC criterion.

The $DF - GLS$ tests is again a t-statistic, yet differs from the ADF test by replacing the OLS detrending data with the GLS one in (4). The detrended GLS series is constructed as

$$\hat{y}_t = y_t - \hat{\beta}^{g\prime} x_t, \quad (5)$$

where given some chosen $\bar{\alpha} = 1 + \bar{c}/T$, $\hat{\beta}^g = \arg\min \sum_{t=1}^{T} (y_t^{\bar{\alpha}} - \beta' x_t^{\bar{\alpha}})^2$, where for any series $\{x_t\}_{t=0}^{T}$, $(x_0^{\bar{\alpha}}, x_t^{\bar{\alpha}})$=$(x_0, (1-\bar{\alpha}L)x_t)$, $t = 1, 2, \cdots, T$. Note that x_t are 1 and $(1, t)'$, respectively, when $m = 0$ and $m = 1$. The value of $\bar{c}$, as recommended by Elliot *et al.*, is chosen to be -7.0 for $m = 0$ and -13.5 for $m = 1$ under which the asymptotic local power function of the $DF - GLS$ test lies close to the Gaussian local power envelope. The $DF - GLS$ test statistic can then obtained by forming a t-ratio for $a_0 = 0$ by running the autoregression with GLS detrended data,

$$\Delta \hat{y}_t = a_0 \hat{y}_{t-1} + \sum_{i=1}^{p} a_i \Delta \hat{y}_{t-i} + e_{tp} \quad (6)$$

[1] However, our analysis that follows can be extended immediately to other test statistics not considered here, or under the GLS detrending, with suitable modifications.

The limiting distributions of the ADF and $DF-GLS$ have been investigated in the literature. We now restate them.

Lemma 1: If $\{y_t\}_1^T$ is generated as in (1) and (2), and is transformed by the GLS local detrending as in (5). Then under the null hypothesis that $\alpha = 1$, as $T \to \infty$, if $m = 0$,

$$DF \Rightarrow 0.5(\tilde{W}^2(1) - 1)\left(\int_0^1 \tilde{W}^2(s)ds\right)^{-1/2} \quad (7)$$

$$DF-GLS \Rightarrow 0.5(W^2(1) - 1)\left(\int_0^1 W^2(s)ds\right)^{-1/2} \quad (8)$$

and if $m = 1$,

$$DF \Rightarrow 0.5(\tilde{W}^2(1) - 1)\left(\int_0^1 \tilde{W}^2(s)ds\right)^{-1/2} \quad (9)$$

$$DF-GLS \Rightarrow 0.5(V^2(1,\bar{c}) - 1)\left(\int_0^1 V^2(s,\bar{c})ds\right)^{-1/2} \quad (10)$$

where W is the standard Wiener process on $[0,1]$. $\tilde{W}(r) = W(r) - \int_0^1 W(s)ds$ is a demeaned Wiener process for $m = 0$, while $\tilde{W}(r) = W(r) + 2\int_0^1 W(s)ds - 6\int_0^1 sW(s)ds$ a detrended Wiener process for $m = 1$. Furthermore, $V(r,\bar{c}) = W(r) - r[\lambda W(1) + 3(1-\lambda)\int_0^1 sW(s)ds]$, and $\lambda = (1-\bar{c})/(1-\bar{c}+\bar{c}^2/3)$.

2.2 MZ and $MZ-GLS$ tests

The MZ tests proposed by Perron and Ng (1996) are based on some modifications to the Z tests. The latter are found to possess considerable size distortions, particularly in the presence of negative moving average errors. The finite-sample size performance of the MZ tests, instead, are shown to be superior to that with the Z tests. There are 3 tests considered in Perron and Ng (1996), MZ_α, MZ_t and MSB. They are defined as

$$(11)\ MZ_\alpha = \left(\frac{\tilde{y}_T^2}{T} - s_{AR}^2\right)\left(\frac{T^2}{2\sum_{t=1}^T \tilde{y}_{t-1}^2}\right)$$

$MSB = \left(\frac{T^{-2}\sum_{t=1}^T \tilde{y}_{t-1}^2}{s_{AR}^2}\right)^{\frac{1}{2}}$ and $MZ_t = MZ_\alpha \times MSB$, where again $\tilde{y}_t$ is the OLS detrended series.[2] Computation of all three tests requires s_{AR}^2, a parametric estimate of the spectral density at frequency zero of v_t. It is then computed as

$$s_{AR}^2 = \frac{\tilde{\sigma}_p^2}{(1 - \sum_{i=1}^p \tilde{a}_i)^2}$$

where $\tilde{a}_i$ and $\tilde{e}_{tp}$ are, respectively, the OLS estimates of coefficients on lagged differenced terms and residuals obtained from running regression in (4), and $\tilde{\sigma}_p^2 = \frac{\sum_{t=p+1}^T \tilde{e}_{tp}^2}{T-p}$. The advantage of s_{AR}^2 over other non-parametric consistent counterparts such as those kernel-based ones lies in its efficiency in estimation of the spectral density at frequency zero, particularly when moving average errors present (see Perron and Ng, 1998). For ease of exposition, we only take up the MZ_α test statistics below, because all three tests share similar asymptotic properties.

To improve the power, it is natural to apply the idea of the GLS detrending to the MZ tests. This is considered by Ng and Perron (2001). Replacing $\tilde{y}_t$ by $\hat{y}_t$ in (11), we define

$$(12)\ MZ_\alpha - GLS = \left(\frac{\hat{y}_T^2}{T} - s_{AR}^2\right)\left(\frac{T^2}{2\sum_{t=1}^T \hat{y}_{t-1}^2}\right)$$

The following lemma summarizes the asymptotic behaviors of the MZ_α and $MZ_\alpha - GLS$ tests.

Lemma 2: Let s_{AR}^2 be a consistent estimate of σ^2. Given conditions stated in *Lemma 1*, as $T \to \infty$, if $m = 0$,

$$MZ_\alpha \Rightarrow 0.5(\tilde{W}^2(1) - 1)\left(\int_0^1 \tilde{W}^2(s)ds\right)^{-1/2} \quad (13)$$

$$MZ_\alpha - GLS \Rightarrow 0.5(W^2(1) - 1)\left(\int_0^1 W^2(s)ds\right)^{-1/2} \quad (14)$$

and if $m = 1$,

$$MZ_\alpha \Rightarrow 0.5(\tilde{W}^2(1) - 1)\left(\int_0^1 \tilde{W}^2(s)ds\right)^{-1/2} \quad (15)$$

$$MZ_\alpha - GLS \Rightarrow 0.5(V^2(1,\bar{c}) - 1)\left(\int_0^1 V^2(s,\bar{c})ds\right)^{-1/2} \quad (16)$$

where $\tilde{W}(r)$, $V(r,\bar{c})$, and λ are as those defined in *Lemma 1*.

[2] Indeed, it is easy to show that $MZ_\alpha = Z_\alpha + (T/2)(\tilde{a}_0 - 1)^2$ where $\tilde{a}_0$ is the OLS estimate from regression (4).

Th result simply shows that $MZ_\alpha - GLS$ test converges to the same limit distribution as that of the $DF - GLS$ test under the null. Although not illustrated here, both tests even have the same asymptotic local behaviors by allowing $\alpha = 1 + c/T$. In other words, these tests achieves the same asymptotic local power that is close to the power envelope.

3 The Bootstrap Resampling Proposal

Implementation of the unit root test statistics as in the preceding discussion entails the asymptotic critical values, typically generated by simulations. The difficulty with the traditional approach to testing for unit root based on the asymptotic critical values, however, is that the nominal size may quite differ from the actual one in the small samples. The distortion deteriorates when negative moving-average errors occur. The bootstrap then serves a reasonable alternative to the conventional asymptotic approximations.

We base our bootstrap resampling scheme on several previous work. The bootstrap procedure proposed by Nankervis and Savin (1996) work for the DF tests with iid errors. The bootstrap considered by Ferreti and Romo (1996) allows for $AR(1)$ errors. Our methods are thus a natural extension of these earlier contributions to allow for weakly dependent errors. We establish the asymptotic validity for most of the available unit root test statistics by showing that the bootstrap test statistics converge to the limiting distributions.

We now detail our bootstrap resampling schemes.

1. Suppose a sample $\{y_t\}_{t=1}^T$ is generated from (1) and (2). Let η denote the nuisance parameters. Depending on whether $m = 0$ or $m = 1$, $\eta = (\phi_1, \cdots, \phi_{p(T)}, F)$ or $\eta = (\beta_0, \phi_1, \cdots, \phi_{p(T)}, F)$. To estimate η,
   (a) compute the residual of constrained least square (CLS); $\hat{v}_t$, by imposing the null hypothesis that $\alpha = 1$. For $m = 0$, $\hat{v}_t = \Delta y_t$, while for $m = 1$, $\hat{v}_t = \Delta y_t - \hat{\beta}_0$, where $\hat{\beta}_0 = T^{-1} \sum_{t=1}^T \Delta y_t$;
   (b) fit an $AR(p)$ to $\hat{v}_t$ where the lag order p is chosen by the AIC or BIC criterion; i.e. $\hat{v}_t = \sum_{i=1}^p \hat{\phi}_i \hat{v}_{t-i} + \hat{\epsilon}_t$;
   (c) center the residual $\{\bar{\epsilon}_t\}$ by

   $$\bar{\epsilon}_t \equiv \hat{\epsilon}_t - \frac{1}{T}\sum_{t=1}^T \hat{\epsilon}_t. \tag{17}$$

   (d) Draw a sample of size T with replacement from the empirical distribution function of $\{\bar{\epsilon}_t\}$, F_T, and denote it by ϵ_t^*.

2. Generate the bootstrap samples $\{y_t^*\}$ according to

$$v_t^* = \sum_{i=1}^p \hat{\phi}_i v_{t-i}^* + \epsilon_t^* \tag{18}$$

$$y_t^* = y_{t-1}^* + v_t^* \quad (m = 0); \quad = \hat{\beta}_0 + y_{t-1}^* + v_t^* \quad (m = \tag{19}$$

3. On the basis of the resample $\{y_t^*\}$, compute the bootstrap counterparts of ADF, $DF-GLS$, MZ_α and $MZ_\alpha-GLS$ as described above, denoted by ADF^*, $DF - GLS^*$, MZ_α^* and $MZ_\alpha - GLS^*$.

4. Repeat step 1 to 3 NB times.

5. Compute the empirical distribution function (edf) of NB values of various bootstrap test statistics under study, and use this empirical distribution function as an approximation to the cumulative distribution function (cdf) of the bootstrap null distribution for the test statistics.

6. Make inference based on the bootstrap critical values.

Some words are worth mentioning. The CLS residual is computed with the constraint of a unit root in series, regardless of whether the sample is from the null or the alternative. It is easy to see how the constraint that $\alpha = 1$ in Step 1.(a) affects the computation of the CLS residual for different models ($m = 0, 1$), given the parameter relation in (3). The importance of this constraint shows again when establishing the asymptotic validity for the bootstrap test statistics later.

An fit of $AR(p)$ model to $\hat{v}_t$ is motivated by an autoregrssive approximation to the infinite-order moving errors in order to reproduce the dependence structure of the data. In which, the lag orders should increase as the sample size increases, i.e. $p = p(T)$. We will give the speed of the lag order to ensure the consistency of the bootstrap approximation. Step 2 is known to be the recursive bootstrap. There are some comparable resampling procedures in the literature, the moving block bootstrap (Künsch, 1989) and stationary bootstrap (Politis and Romano, 1994). In contrast to the latter two procedures, the recursive bootstrap makes use of the $ARMA$ parametric model in reproducing error dependence. Bühlmann (1997) and Horowitz (2000) both emphasize the merit of the use of the recursive bootstrap.

Centering of the residual in Step 1.(c) not only takes into account that the underlying population distribution has zero expectation, but also works to reduce the downward bias of the autoregression coefficients in small samples (see Horowitz, 2000).

To generate the bootstrap resample $\{y_t^*\}$, we need first to choose starting values $(y_1^*, \cdots, y_p^*)$. An inappropriate choice of the initials could cause systematic finite-sample bias. To avoid these shortcoming, we choose $y_i^* = \epsilon_i^*$ for $i = 1, \cdots, p$, where ϵ_i^* are iid drawn from $\bar{F}_T$ and then start generating y_t^* (see Swanepoel and van Wyk, 1986).

4 Bootstrap Consistency

Whether the bootstrap distributions of the test statistics can be justified asymptotically, i.e the consistency of the bootstrap distribution, is an important issue for the bootstrap testing. This is because the bootstrap distribution is to approximate the unknown small-sample distribution, and thus is expected to be asymptotically equivalent to the limit distribution, if the bootstrap procedure is a reasonable one.

Our main assumptions for the bootstrap to yield consistency are the following.

Assumption A: $v_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$, where $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$, $E|\epsilon_t|^4 < \infty$, $\psi_0 = 1$, $\Psi(L)$ is non-zero on unit circle, and $\sum_{i=0}^{\infty} i|\psi_i| < \infty$.

We now present the first set of the theoretical results regarding the consistency for the bootstrap parameters.

Theorem 1 Let *Assumption A* with $p(T) = o((T/\log(T))^{1/4})$ hold. Then, conditional on the sample, as $T \to \infty$,

1. when $m = 1$,

$$T(Var^*(\sum_{t=1}^{T} \frac{\Delta y_t^*}{T}) - Var(\sum_{t=1}^{T} \frac{\Delta y_t}{T})) = o(1).$$

2.

$$\frac{\tilde{\sigma}_p^{*2}}{(1 - \sum_{i=1}^{p} \tilde{a}_i^*)^2} - s_{AR}^2 = o_p(1) \text{ almost surely}$$

where $\tilde{a}_i^*$ and $\tilde{\sigma}_p^{*2} = \frac{\sum_{t=p+1}^{T} \tilde{e}_{tp}^{*2}}{T-p}$ are OLS estimates obtained by running $\Delta\tilde{y}_t^* = a_0 \tilde{y}_{t-1}^* + \sum_{i=1}^{p(T)} a_i \Delta\tilde{y}_{t-i}^* + e_{tp}^*$.

Theorem 1.1 shows the consistency for the bootstrap intercept in the ADF regression (3) by establishing the consistency of second moments. The consistency is obtained under the condition that the autoregressive lag order in the bootstrap ADF regression should grow as the sample increases. The condition for the lag order however is weaker than that specified by Said and Dicky (1984) or Ng and Perron (1996, JASA). In practice, this suggests a smaller number of lag to be fit in the bootstrap ADF regression. The estimate for the bootstrap mean is required in the process of detrending the bootstrap resamples $\{y_t^*\}$. Of significance from the result is that the bootstrap intercept can be consistently estimated, regardless of whether the observed data is drawn from the null or from the alternative. On the other hand, the estimate for the bootstrap autoregressive spectral density at frequency zero is not asymptotically distinguished from the conventional the autoregressive spectral estimate. The consistency guarantees the bootstrap unit root statistics free of the nuisance parameters in the limit. It should be emphasized that with additional assumptions, we are also able to establish the consistency for the distribution function of the bootstrapped parameters. The behavior is not of major concern, and thus is omitted here.

To derive the asymptotic distributions of the bootstrap unit root statistics, it necessitates a functional central limit theory for the bootstrap resamples. We adopt the Mallow metric which is defined as

$$d_2(x, y) = \inf E(\|X - Y\|^2)^{1/2}$$

where X and Y are random variables with corresponding distribution x and y. Note that the fact that $d_2 \rightarrow 0$ implies a convergence in distribution, a crucial step in establishing the asymptotic validity of the bootstrap distribution. Bickel and Freedman (1981) elaborate the statistical properties of the Mallow metric. The following is to provide the invariance principle for the bootstrap partial sum process.

Theorem 3: For any sample, suppose $\{\epsilon_t^*\}$ is reproduced according to the preceding bootstrap algorithm defined in (d). Letting $T \rightarrow \infty$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \epsilon_t^* \quad \Rightarrow \quad \sigma_\epsilon W(r)$$

With the bootstrap invariance principle, we are in a position to demonstrate the consistency of the bootstrap distribution of the unit root statistics.

Theorem 4: Let *Assumption A* with $p(T) = o((T/\log(T))^{1/4})$ hold. Then, conditional on the sample, for any $x \in R$ and $m = 0$ and 1,

1.

$$\sup |P^*(ADF^* \leq x) - P(ADF \leq x|H_0)| = o_p(1)$$

2.

$$\sup |P^*(DF\text{–}GLS^* \leq x) - P(DF\text{–}GLS \leq x|H_0)| =$$

It is important to note that this consistency result stands, whether or not the data is sampled from the null or from the alternative. This property works to equip the bootstrap tests with power when the truth is from the alternative. If the proposed bootstrap fails to replicate the null distribution but the alternative distribution under H_a, the bootstrap tests will possess little power. While there is such a small amount of the literature that has established the asymptotic validity for some bootstrap unit root statistics, our result appears to be the first that proves the consistency to hold even under the trend stationary alternative hypothesis. Of more significance is that the consistency holds with weakly dependent errors. In contrast, the consistency shown by Ferretti and Romo (1996) is valid for the independent and autoregressive errors. Also most of the literature focuses on the simple autoregressive model without deterministic trend components. In the case of their presence in the model as we consider here, making the consistency argument proves to be not quite a straightforward extension where the existence of the consistency of the bootstrap trend is a prerequisite for that of the bootstrap distribution.

The asymptotic validity indeed comes from imposing the unit root when generating the bootstrap resamples, as emphasized in the above algorithm. Basawa *et al.* (1991a,b) show that the bootstrap least square estimator does not converge to the correct asymptotic unit root process, unless the unit root constraint is placed in the resampling.

The bootstrap procedures also work for the class of MZ test statistics as the following theorem states.

Theorem 5: Again, let *Assumption A* with $p(T) = o((T/\log(T))^{1/4})$ hold. Then, conditional on the sample, for any $x \in R$ and $m = 0$ and 1,

1.

$$\sup |P^*(MZ^* \leq x) - P(MZ \leq x|H_0)| = o_p(1)$$

2.

$$\sup |P^*(MZ\text{–}GLS^* \leq x) - P(MZ\text{–}GLS \leq x | H_0)| = o_p(1).$$

We nevertheless make no claim of the asymptotic refinement for our bootstrap proposal. This appears very difficult because it remains unknown whether the Edgeworth expansion is available for the model under study. Such an expansion is only known to exist for the simple Gaussian random-walk model, and none beyond the case (see Abadir, 1993). Given the complicated nature of our model, it may be very likely that the asymptotic refinement is impossible, based on the current knowledge. We leave it for future research.

5 Monte Carlo Simulations

To assess the finite-sample performance of the bootstrap test statistics, we report two sets of experiments under different error processes, $AR(1)$ and $MA(1)$. Because we are using the autoregression approximation in reproducing the samples, it is important to have a correct selection of lag length. This is particularly important in the presence of moving-average errors. Ng and Perron (2001) in fact are in an attempt to better choose the lag by modifying the information criterion in the construction of the unit root tests. Our theory is silent on how exact lags should be chosen for any particular small samples, but reveals that the lags increase with sample sizes. Following the practice, we set a maximum lag orders (5 lags in the simulations), and employ both the AIC and BIC criterion to select corresponding appropriate lags for the various tests under study.

The data generating process is as described in the notes to the tables. The simulation results are reported from Table 1 to 8. The replication for the asymptotic tests is 5,000, and for the bootstrap counterparts 1,000. The sample sizes considered are 50, 100, 150 and 300.

There are a number of conclusions emerging from the simulations. First, our bootstrap schemes appears to be able to well control the size of the tests. Indeed for even the very small sample sizes of 50, the empirical rejection frequencies for the bootstrap tests is still close to the nominal level (5%). The performance of the bootstrap tests stands well still, regardless of the error dependence structure, either MA or AR. This is considered to be remarkable, given that the literature has documented the poor performance of the tests in the case of MA errors. In contrast, the size performance of the asymptotic tests, with different version of modification by construction, comes to be reasonable when the sample sizes is near 100 or larger. Next, the bootstrap tests have the comparable power as the asymptotic counterparts. Note that the power of the asymptotic tests is obtained adjusted for the size distortion. In other words, the power reported for the asymptotic tests is infeasible, because the finite-sample critical values are generally unknown in applications. In some instances, the bootstrap tests show a power gain over the asymptotic counterparts to some minor extent. For example, for most of sample sizes with MA errors, our bootstrap tests is 10% more in power than the asymptotic tests. While we have proven only the bootstrap consistency under the null, the power of the bootstrap tests increases as the sample size grows. This demonstrates, instead, a reflection of another aspect of the consistency for the bootstrap tests under the alternative. Overall, our bootstrap resampling procedures show a better size control at no expense of power loss, comparing to the asymptotic tests.

6 Conclusions

This paper proposes the bootstrap procedures for the presence of a unit root in an AR model. Our schemes are useful in a wide range of applications because it is applicable for most of dependent error structure, while the previous proposals are only valid in the case of iid error. Our bootstrap procedures are shown to give the commonly-used unit root tests excellent performance, in particular the good control over the size in the very small sample sizes. We also offer asymptotic justification for the bootstrap proposals by demonstrating their consistency. While the bootstrap tests are able to reduce the size distortion, there still exists room for making improvement when the MA coefficient is close to the unity region. Another set of our simulations (not reported) show that in this case, our bootstrap tests does not perform as well as it does when the MA coefficient is away from the unity region, in reducing the size distortion. This problem is not unique to the bootstrap tests. The asymptotic tests we consider have a much worse-off size distortion in the case. Future research is called for an alternative bootstrap that still yields reasonable size control in this problematic region.

7 Appendix

If $m = 1$, the series is generated by

$$\begin{aligned} y_t &= \beta_0 + \beta_1 t + u_t; \\ u_t &= \alpha u_{t-1} + v_t; \end{aligned}$$

where $v_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, $\psi_0 = 1$, $\sum_{j=0}^{\infty} j|\psi_j| < \infty$, $\varepsilon_t \overset{iid}{\;} (0, \sigma_\varepsilon^2)$, $E|\varepsilon_t|^4 < \infty$.

Under $H_0 : \alpha = 1$,

$$\triangle y_t = \beta_1 + v_t, \quad v_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

Because $v_t \sim AR(\infty)$, it can be re-written:

$$\sum_{j=0}^{\infty} \phi_j v_{t-j} = \varepsilon_t, \quad \phi_0 = 1.$$

Thus, we shall exploit $\hat{v}_t = \triangle y_t - \hat{\beta}_1$ to estimate $AR(p)$, where $\hat{\beta}_1$ is the OLS estimate of β_1 and $p = o((T/\lg T)^{1/4}$. In addition, denote $\hat{\phi}_{1,T}, \hat{\phi}_{2,T}, \cdots, \hat{\phi}_{p,T}$ as the OLS estimates of $\phi_1, \phi_2, \cdots, \phi_p$, equal to those obtained by the Yule-Walker regression.

As a consequence,

$$(A1) \qquad \sum_{i=0}^{p(T)} \hat{\phi}_{i,T} \hat{v}_{t-i} = \hat{\varepsilon}_{t,T}$$

where $\{\hat{\varepsilon}_{t,T}\}_{t=p+1}^{T}$ are the residuals. Let $\bar{\varepsilon}_T = (T-p)^{-1} \sum_{t=p+1}^{T} \hat{\varepsilon}_{t,T}$, $\tilde{\varepsilon}_{t,T} = \hat{\varepsilon}_{t,T} - \bar{\varepsilon}_T$. Draw samples from $\{\tilde{\varepsilon}_{t,T}\}_{t=p+1}^{T}$, denoted by $\{\varepsilon_t^*\}$. Reproducing the error process based on (A1), we have:

$$(A2) \qquad \sum_{i=0}^{p(T)} \hat{\phi}_{i,T} \hat{v}_{t-i}^* = \hat{\varepsilon}_{t,T}^*.$$

Then reproduce the series by :

$$\triangle y_t^* = \hat{\beta}_1 + v_t^*,$$

denoted by $\{y_t^*\}_{t=1}^{T}$.

Let $\hat{\Phi}_T(L) = \sum_{i=0}^{\infty} \hat{\phi}_{i,T} L^i$, and $\hat{\Psi}_T(L) = \frac{1}{\hat{\Phi}_T(L)} = \sum_{i=0}^{\infty} \hat{\psi}_{i,T} L^i$. Given the assumptions about DGP and $p(T)$, there exists some T_1 such that

$$\sup_{T > T_1} \sum_{i=0}^{\infty} j|\hat{\psi}_{j,T}| < \infty, \quad \sup_{0 \le j < \infty} |\hat{\psi}_{j,T} - \psi_j| = o(1).$$

Denote $\hat{F}_{\varepsilon,T}$ to be the empirical distribution function (edf) of $\{\tilde{\varepsilon}_{t,T}\}_{t=p+1}^{T}$, $F_{\varepsilon,T}$ is the edf of $\{\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_T\}$, F_ε is the edf of $\{\varepsilon_1, \varepsilon_2, \cdots, \}$. As a result,

$$d_2(\hat{F}_{\varepsilon,T}, F_\varepsilon) \le d_2(\hat{F}_{\varepsilon,T}, F_{\varepsilon,n}) + d_2(F_{\varepsilon,T}, F_\varepsilon)$$

Due to Bickel and Freedman (1981), $d_2(F_{\varepsilon,T}, F_\varepsilon) \longrightarrow 0$. Also, by Bühlmann (1997), $d_2(\hat{F}_{\varepsilon,T}, F_{\varepsilon,T}) \longrightarrow 0$, we have

$$d_2(\hat{F}_{\varepsilon,T}, F_\varepsilon) \longrightarrow 0,$$

that is,

$$\varepsilon_t^* \xrightarrow{d^*} \varepsilon_t.$$

Because $d_2(\hat{F}_{\varepsilon,T}, F_\varepsilon) \longrightarrow 0$, and there exist $\{\varepsilon_t\}$ and $\{\varepsilon_t^*\}$ such that $\inf d_2$ holds,

$$d_2(\hat{F}_{\varepsilon,T}, F_\varepsilon) = (E|\varepsilon_t - \varepsilon_t^*|^2)^{1/2}$$

Let $S_{[Tr]}$ and $S_{[Tr]}^*$ be defined as:

$$\begin{aligned} S_{[Tr]} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \varepsilon_t, \\ S_{[Tr]}^* &= \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \varepsilon_t^* \end{aligned}$$

and based on FCLT, $S_{[Tr]} \Longrightarrow \sigma_\varepsilon W(r)$.

Let $V_{n,k}$ and $V^*_{n,k}$ be given to be:

$$\begin{aligned} V_{T,k} &= [S_{[Tr_1]}, S_{[Tr_2]} - S_{[Tr_1]}, \cdots, S_{[Tr_k]} - S_{[Tr_{k-1}]}]' \\ V^*_{T,k} &= [S^*_{[Tr_1]}, S^*_{[Tr_2]} - S^*_{[Tr_1]}, \cdots, S^*_{[Tr_k]} - S^*_{[Tr_{k-1}]}]' \end{aligned}$$

where $0 < r_1 < r_2 < \cdots < r_k$. Therefore,

$$\begin{aligned} d_2^2(V_{T,k}, V^*_{T,k}) &\leq \frac{1}{T} \sum_{j=1}^{k} \sum_{t=[Tr_{j-1}]+1}^{[Tr_j]} E(\varepsilon_t - \varepsilon^*_t)^2 \\ &= \frac{1}{T} \sum_{t=1}^{[Tk]} E(\varepsilon_t - \varepsilon^*_t)^2 \\ &\leq d_2^2(\hat{F}_{\varepsilon,T}, F_\varepsilon) \\ &\longrightarrow 0 \end{aligned}$$

In other words, $S^*_{[Tr]}$ converges to $S^*_{[Tr]}$ in finite-dimension distribution.

Now to prove the tightness of $S^*_{[Tr]}$. For $0 \leq r_1 \leq r \leq r_2$, there exists a non-decreasing function such that

$$P^*(|S^*_{[Tr]} - |S^*_{[Tr_1]}| \geq \lambda, |S^*_{[Tr_2]} - |S^*_{[Tr]}| \geq \lambda) \leq \lambda^{-4}(\varphi(r_2) - \varphi(r_1))^2$$

This holds because

$$\begin{aligned} P^*(|S^*_{[Tr]} - |S^*_{[Tr_1]}| \geq \lambda, |S^*_{[Tr_2]} - |S^*_{[Tr]}| \geq \lambda) &\leq \lambda^{-4} E|S^*_{[Tr]} - S^*_{[Tr_1]}|^2 * E|S^*_{[Tr_2]} - |S^*_{[Tr]}|^2 \\ &\equiv \lambda^{-4} A \end{aligned}$$

where by letting $\varphi(x) = xE(\varepsilon^{*2}_t)^2$

$$\begin{aligned} A &= E|S^*_{[Tr]} - S^*_{[Tr_1]}|^2 * E|S^*_{[Tr_2]} - |S^*_{[Tr]}|^2 \\ &= \frac{1}{T} \sum_{j=[Tr_1]+1}^{[Tr]} E(\varepsilon^{*2}_t) * \frac{1}{T} \sum_{j=[Tr]+1}^{[Tr_2]} E(\varepsilon^{*2}_t) \\ &\leq (r_2 - r_1)^2 E(\varepsilon^{*2}_t)^2 \\ &= (\varphi(r_2) - \varphi(r_1))^2. \end{aligned}$$

This proves the tightness of $S^*_{[Tr]}$, giving

$$S^*_{[Tr]} \Longrightarrow \sigma_\varepsilon W(r)$$

Recall that $v_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \Psi(L)\varepsilon_t$ and $\sum_{j=0}^{\infty} j|\psi_j| < \infty$. We thus have $V_{[Tr]} = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} v_t \Longrightarrow \omega W(r)$, $\omega = \Psi(1)\sigma_\varepsilon$.

Because $v^*_t = \sum_{j=0}^{\infty} \hat{\psi}_{j,T} \varepsilon^*_{t-j} = \hat{\Psi}(L)\varepsilon^*_t$, and let $V^*_{[Tr]} = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} v^*_t$, we need to show:

$$V^*_{[Tr]} \Longrightarrow \omega W(r)$$

Let $v^{**}_t = \sum_{j=0}^{\infty} \psi_j \varepsilon^*_t$, $V^{**}_{[Tr]} = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} v^{**}_t$. Because $V^{**}_{[Tr]} \Longrightarrow \omega W(r)$, the claim above would hold provided that

$$\sup_{r \in [0,1]} |V^*_{[Tr]} - V^{**}_{[Tr]}| \xrightarrow{p} 0$$

This can be shown to hold as follows. Given any arbitrary $r \in [0,1]$, and $\eta > 0$,

$$P^*(|V^*_{[Tr]} - V^{**}_{[Tr]}| > \eta) < \frac{E|V^*_{[Tr]} - V^{**}_{[Tr]}|}{\eta}$$

where

$$\begin{aligned} E|V^*_{[Tr]} - V^{**}_{[Tr]}| &= E|\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} (\hat{\Psi}(L) - \Psi(L))\varepsilon^*_t| \\ &\leq \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} |\hat{\Psi}(L) - \Psi(L)| E|\varepsilon^*_t| \\ &= E|\varepsilon^*_t| \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} (\sum_{j=0}^{\infty} |\hat{\psi}_j - \psi_j|) \\ &= rE|\varepsilon^*_t| \sum_{j=0}^{\infty} |\hat{\psi}_j - \psi_j| \\ &\longrightarrow 0. \end{aligned}$$

This is because $\sup_{0 \leq j < \infty} |\hat{\psi}_j - \psi_j| \longrightarrow 0$ as $T \longrightarrow \infty$. So, $\forall r \in [0,1]$, $|V^*_{[Tr]} - V^{**}_{[Tr]}| \xrightarrow{p} 0$,

$$\sup_{r \in [0,1]} |V^*_{[Tr]} - V^{**}_{[Tr]}| \xrightarrow{p} 0.$$

We thus have reached the following important result,

$$V^*_{[nr]} \Longrightarrow \omega W(r).$$

Denote $\tilde{y}^*_t$ to be the GLS detrending y^*_t. Because $y^*_t = \hat{\beta}_1 + y^*_{t-1} + v^*_t$, by the same argument as in Elliot *et al.* (1996), we obtain:

$$\frac{1}{\sqrt{T}} \tilde{y}^*_{[Tr]} \Longrightarrow \omega V_c(r, \bar{c})$$

Because $DF - GLS^*$ is a t-statistic for $H_0 : a_0 = 0$ in the following regression,

$$\triangle \tilde{y}^*_t = a_0 + a_1 \tilde{y}^*_{t-1} + \triangle \tilde{y}^*_{t-1} + \cdots + \triangle \tilde{y}^*_{t-p} + e_{tp}$$

As deriving the asymptotics of $DF - GLS$, we finally have:

$$DF - GLS^* \Longrightarrow \frac{V_c(1, \bar{c}) - 1}{2(\int_0^1 V_c^2(r, \bar{c})^{1/2}}$$

which have the same limiting distribution as $DF - GLS$ as asserted. Similar arguments are applicable to ADF^*, MZ^* and $MZ - GLS^*$.

References

[1] Abadir, K.M. (1993), The limiting distribution of the autocorrelation coefficient under a unit root, *Annals of Statistics*, 21, 1058-1070.

[2] Andrews, D.W.K. and M. Buchinsky (1999), On the number of bootstrap repetitions for bootstrap standard errors, Confidence Intervals, and Tests, *Econometrica,* forthcoming.

[3] Basawa, I. V., A. K. Mallik, W. P. McCormick and R. L. Taylor (1989), Bootstrapping explosive autoregressive process, *Annals of Statistics,* 17, 1479-1486.

[4] Basawa, I. V., A. K. Mallik, W. P. McCormick and R. L. Taylor (1991a), Bootstrapping unstable first order autoregressive process, *Annals of Statistics,* 19, 1098-1101.

[5] Basawa, I. V., A. K. Mallik, W. P. McCormick and R. L. Taylor (1991b), Bootstrapping test of significance and sequential bootstrap estimation for unstable first order autoregressive process, *Commun. Stat. Theory Methods,* 20, 1015-1026.

[6] Berk, K.N. (1974), Consistent autoregressive spectral estimates, *Annals of Statistics,* 2, 489-502.

[7] Bickel, P. and D. Freedman (1981), Some asymptotic theory for the bootstrap, *Annals of Statistics,* 9, 1196-1217.

[8] Billingsley, P. (1968), Convergence of probability measures, New York: Wiley.

[9] Bühlmann, P. (1997), Sieve bootstrap for time series, *Bernoulli,* 3, 123-148.

[10] 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.

[11] Elliott, G., T. J. Rothenberg and J. H. Stock (1996), Efficient tests for an autoregressive unit root, *Econometrica,* 64, 813-836.

[12] Ferretti, N. and J. Romo (1995), Bootstrapping unit root AR(1) models, *Biometrika,* 82, 849-860.

[13] Horowitz, L. (2000), The bootstrap, *Handbook of Econometrics,* Vol.4, Amesterdam: North-Holland, forthcoming.

[14] Künsch, H. R. (1989), The jackknife and the bootstrap for general stationay observations, *Annals of Statistics,* 17, 1217-1241.

[15] Li, H. and G. S. Maddala (1996), Bootstrapping time series models, *Econometric Reviews,* 15, 115-158.

[16] Li, H. and G. S. Maddala (1997), Bootstrapping cointegrating regressions, *Journal of Econometrics,* 80, 297-318.

[17] Nankervis, J. and N. Savin (1996), The level and power of the bootstrap t test in the AR(1) model with trend, *Journal of Business & Economic Statistics,* 14, 161-168.

[18] Ng, S. and P. Perron (1995), Unit root tests in ARMA models with data-dependent methods for the selection of truncation lag, *Journal of the American Statistical Association,* 90, 268-281.

[19] Ng, S. and P. Perron (1998), An autoregressive spectral density estimator at frequency zero For nonstationary tests, 14/5, 560-603.

[20] Ng, S. and P. Perron (2001), Lag length selection and the construction of unit root tests with good size and power, *Econometrica,* forthcoming.

[21] Phillips, P. C. B. and P. Perron (1988), Testing for a unit root in a time series regression, *Biometrika,* 75, 335-346.

[22] Phillips, P. C. B. and V. Solo (1992), Asymptotics for linear processes, *Annals of Statistics,* 20, 971-1001.

[23] Phillips, P. C. B. (2001), Bootstrapping spurious regression, Cowles Foundation Discussion Paper No. 1330, Yale University.

[24] Perron, P. C. B. and S. Ng (1996), Useful modifications to some unit root tests with dependent errors and their local asymptotic properties, *Review of Economic Studies,* 63, 435-465.

[25] Said, S.E. and D.A. Dicky (1984), Tesitng for unit root in time series regression, *Biometrika*, 75, 599-608.

Table 1: Size and Power of Unit Root Test Statistics, $DF-GLS$

Panel A: $m=0$ and AR(1) error

T	α	$DF-GLS$	$DF-GLS^*$	$DF-GLS$	$DF-GL$
		AIC	AIC	BIC	BIC
50	1	0.128	0.040	0.113	0.057
100	1	0.082	0.042	0.075	0.050
150	1	0.069	0.050	0.066	0.056
300	1	0.056	0.052	0.054	0.050
50	0.90	0.219	0.165	0.223	0.197
100	0.90	0.556	0.436	0.585	0.536
150	0.90	0.816	0.722	0.848	0.794
300	0.90	0.998	0.989	0.999	0.998
50	0.85	0.314	0.231	0.334	0.293
100	0.85	0.759	0.616	0.803	0.737
150	0.85	0.944	0.863	0.970	0.945
300	0.85	1.000	0.999	1.000	1.000
50	0.80	0.406	0.293	0.439	0.367
100	0.80	0.868	0.736	0.913	0.859
150	0.80	0.978	0.934	0.994	0.981
300	0.80	1.000	1.000	1.000	1.000

Panel B: $m=1$ and AR(1) error

T	α	$DF-GLS$	$DF-GLS^*$	$DF-GLS$	$DF-GL$
		AIC	AIC	BIC	BIC
50	1	0.164	0.027	0.129	0.022
100	1	0.094	0.036	0.078	0.041
150	1	0.073	0.042	0.065	0.054
300	1	0.052	0.038	0.047	0.043
50	0.90	0.085	0.040	0.090	0.043
100	0.90	0.211	0.137	0.225	0.217
150	0.90	0.423	0.339	0.456	0.401
300	0.90	0.947	0.866	0.969	0.936
50	0.85	0.116	0.054	0.128	0.074
100	0.85	0.360	0.236	0.396	0.371
150	0.85	0.690	0.532	0.742	0.650
300	0.85	0.995	0.973	0.999	0.994
50	0.80	0.151	0.075	0.175	0.107
100	0.80	0.516	0.358	0.570	0.517
150	0.80	0.846	0.693	0.898	0.830
300	0.80	0.998	0.994	1.000	0.999

Note:

1. DGP is as described in (1) and (2) where we set $\beta_0=\beta_1=0$. $v_t=\phi v_{t-1}+\epsilon_{t-1}$, with $\phi=0.5$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, and $\alpha=1,0.9,0.85,0.8$.
2. The replications for the asymptotic tests are 5000, while for the bootstrap tests are 1000 (=NB).
3. The 5%-level asymptotic critical value of the $DF-GLS$ test is -1.98 for $m=0$ and -2.91 for $m=1$.
4. The lag orders p fit to the bootstrap ADF regression and to the bootstrap autoregression in reproducing the samples are separately chosen by both AIC and BIC. The maximum lag orders are set to be 5.

Table 2: Size and Power of Unit Root Test Statistics, DF

Panel A: $m = 0$ and AR error

T	α	DF AIC	DF^* AIC	DF BIC	DF^* BIC
50	1	0.087	0.026	0.070	0.037
100	1	0.069	0.036	0.060	0.036
150	1	0.061	0.052	0.056	0.045
300	1	0.055	0.035	0.052	0.042
50	0.90	0.117	0.061	0.116	0.073
100	0.90	0.239	0.165	0.244	0.221
150	0.90	0.449	0.354	0.464	0.422
300	0.90	0.949	0.869	0.969	0.947
50	0.85	0.164	0.077	0.167	0.107
100	0.85	0.407	0.283	0.436	0.388
150	0.85	0.720	0.561	0.756	0.687
300	0.85	0.995	0.982	0.999	0.996
50	0.80	0.214	0.095	0.221	0.136
100	0.80	0.570	0.392	0.616	0.526
150	0.80	0.875	0.717	0.914	0.858
300	0.80	0.999	0.998	1.000	1.000

Panel B: $m = 1$ and AR error

T	α	DF AIC	DF^* AIC	DF BIC	DF^* BIC
50	1	0.130	0.021	0.098	0.019
100	1	0.087	0.028	0.070	0.042
150	1	0.071	0.040	0.062	0.046
300	1	0.059	0.040	0.052	0.045
50	0.90	0.082	0.026	0.079	0.034
100	0.90	0.147	0.091	0.149	0.128
150	0.90	0.283	0.196	0.289	0.257
300	0.90	0.817	0.672	0.855	0.800
50	0.85	0.102	0.038	0.103	0.052
100	0.85	0.241	0.139	0.257	0.222
150	0.85	0.498	0.335	0.525	0.454
300	0.85	0.971	0.894	0.988	0.968
50	0.80	0.126	0.040	0.133	0.066
100	0.80	0.360	0.200	0.385	0.342
150	0.80	0.692	0.472	0.741	0.646
300	0.80	0.994	0.967	0.999	0.988

Note:

1. DGP is as described in (1) and (2) where we set $\beta_0 = \beta_1 = 0$. $v_t = \phi v_{t-1} + \epsilon_{t-1}$, with $\phi = 0.5$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, and $\alpha = 1, 0.9, 0.85, 0.8$.
2. The replications for the asymptotic tests are 5000, while for the bootstrap tests are 1000 (=NB).
3. The 5%-level asymptotic critical value of the DF test is -2.86 for $m = 0$ and -3.41 for $m = 1$.
4. The lag orders p fit to the bootstrap ADF regression and to the bootstrap autoregression in reproducing the samples are separately chosen by both AIC and BIC. The maximum lag orders are set to be 5.

Table 3: Size and Power of Unit Root Statistics, $DF-GLS$

Panel A: $m = 0$ and MA(1) error

T	α	$DF-GLS$ AIC	$DF-GLS^*$ AIC	$DF-GLS$ BIC	$DF-GL$ BIC
50	1	0.238	0.061	0.285	0.108
100	1	0.147	0.067	0.205	0.116
150	1	0.114	0.073	0.174	0.116
300	1	0.081	0.056	0.122	0.078
50	0.90	0.184	0.177	0.196	0.332
100	0.90	0.402	0.427	0.443	0.601
150	0.90	0.594	0.652	0.646	0.748
300	0.90	0.865	0.865	0.885	0.909
50	0.85	0.273	0.245	0.294	0.409
100	0.85	0.540	0.531	0.605	0.679
150	0.85	0.691	0.725	0.752	0.810
300	0.85	0.873	0.878	0.894	0.914
50	0.80	0.366	0.316	0.405	0.490
100	0.80	0.603	0.593	0.688	0.702
150	0.80	0.711	0.728	0.780	0.813
300	0.80	0.863	0.878	0.884	0.909

Panel B: $m = 1$ and MA(1) error

T	α	$DF-GLS$ AIC	$DF-GLS^*$ AIC	$DF-GLS$ BIC	$DF-GL$ BIC
50	1	0.328	0.041	0.373	0.098
100	1	0.217	0.060	0.304	0.142
150	1	0.161	0.060	0.263	0.130
300	1	0.096	0.068	0.172	0.115
50	0.90	0.104	0.079	0.105	0.164
100	0.90	0.227	0.217	0.248	0.389
150	0.90	0.385	0.429	0.443	0.616
300	0.90	0.854	0.867	0.865	0.926
50	0.85	0.155	0.096	0.158	0.213
100	0.85	0.391	0.327	0.440	0.555
150	0.85	0.594	0.607	0.690	0.754
300	0.85	0.922	0.928	0.936	0.961
50	0.80	0.220	0.124	0.229	0.280
100	0.80	0.534	0.439	0.620	0.638
150	0.80	0.706	0.743	0.813	0.837
300	0.80	0.933	0.942	0.949	0.962

Note:

1. DGP is as described in (1) and (2) where we set $\beta_0 = \beta_1 = 0$. $v_t = \epsilon_t + \theta\epsilon_{t-1}$, with $\theta = -0.5$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, and $\alpha = 1, 0.9, 0.85, 0.8$.
2. The replications for the asymptotic tests are 5000, while for the bootstrap tests are 1000 (=NB).
3. The 5%-level asymptotic critical value of the $DF-GLS$ test is -1.98 for $m = 0$ and -2.91 for $m = 1$.
4. The lag orders p fit to the bootstrap ADF regression and to the bootstrap autoregression in reproducing the samples are separately chosen by both AIC and BIC. The maximum lag orders are set to be 5.

Table 4: Size and Power of Unit Root Statistics, DF

Panel A: $m = 0$ and MA(1) error

T	α	DF	DF^*	DF	DF^*
		AIC	AIC	BIC	BIC
50	1	0.181	0.043	0.203	0.077
100	1	0.137	0.066	0.188	0.092
150	1	0.116	0.069	0.176	0.118
300	1	0.077	0.060	0.121	0.081
50	0.90	0.154	0.102	0.164	0.205
100	0.90	0.385	0.282	0.418	0.472
150	0.90	0.575	0.541	0.656	0.716
300	0.90	0.972	0.966	0.981	0.989
50	0.85	0.251	0.143	0.270	0.290
100	0.85	0.619	0.469	0.694	0.668
150	0.85	0.820	0.789	0.889	0.875
300	0.85	0.999	0.997	0.999	0.999
50	0.80	0.377	0.212	0.407	0.393
100	0.80	0.788	0.621	0.867	0.775
150	0.80	0.935	0.920	0.969	0.958
300	0.80	1.000	1.000	1.000	1.000

Panel B: $m = 1$ and MA(1) error

T	α	DF	DF^*	DF	DF^*
		AIC	AIC	BIC	BIC
50	1	0.287	0.032	0.309	0.091
100	1	0.220	0.050	0.295	0.126
150	1	0.176	0.058	0.279	0.115
300	1	0.107	0.068	0.193	0.107
50	0.90	0.098	0.055	0.099	0.134
100	0.90	0.213	0.170	0.224	0.327
150	0.90	0.374	0.331	0.422	0.519
300	0.90	0.857	0.863	0.872	0.926
50	0.85	0.143	0.083	0.145	0.169
100	0.85	0.399	0.283	0.436	0.485
150	0.85	0.647	0.572	0.738	0.737
300	0.85	0.981	0.990	0.990	0.996
50	0.80	0.209	0.107	0.216	0.221
100	0.80	0.596	0.428	0.670	0.610
150	0.80	0.816	0.763	0.903	0.869
300	0.80	0.998	0.999	0.999	0.998

Note:

1. DGP is as described in (1) and (2) where we set $\beta_0 = \beta_1 = 0$. $v_t = \epsilon_t + \theta\epsilon_{t-1}$, with $\theta = -0.5$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, and $\alpha = 1, 0.9, 0.85, 0.8$.
2. The replications for the asymptotic tests are 5000, while for the bootstrap tests are 1000 (=NB).
3. The 5%-level asymptotic critical value of the DF test is -2.86 for $m = 0$ and -3.41 for $m = 1$.
4. The lag orders p fit to the bootstrap ADF regression and to the bootstrap autoregression in reproducing the samples are separately chosen by both AIC and BIC. The maximum lag orders are set to be 5.

Table 5: Size and Power of Unit Root Statistics, $MZ - GLS$

Panel A: $m = 0$ and AR(1) error

T	α	$MZ - GLS$	$MZ - GLS^*$	$MZ - GLS$	$MZ - G$
		AIC	AIC	BIC	BIC
50	1	0.174	0.038	0.139	0.038
100	1	0.101	0.039	0.087	0.047
150	1	0.081	0.050	0.074	0.054
300	1	0.065	0.053	0.060	0.049
50	0.90	0.137	0.125	0.180	0.174
100	0.90	0.477	0.397	0.552	0.520
150	0.90	0.796	0.699	0.837	0.785
300	0.90	0.998	0.990	1.000	0.998
50	0.85	0.182	0.179	0.266	0.260
100	0.85	0.687	0.584	0.774	0.727
150	0.85	0.937	0.857	0.967	0.942
300	0.85	1.000	0.998	1.000	1.000
50	0.80	0.223	0.245	0.344	0.328
100	0.80	0.816	0.709	0.890	0.840
150	0.80	0.972	0.934	0.993	0.979
300	0.80	1.000	1.000	1.000	1.000

Panel B: $m = 1$ and AR(1) error

T	α	$MZ - GLS$	$MZ - GLS^*$	$MZ - GLS$	$MZ - G$
		AIC	AIC	BIC	BIC
50	1	0.261	0.023	0.159	0.028
100	1	0.128	0.034	0.084	0.041
150	1	0.089	0.044	0.067	0.054
300	1	0.058	0.037	0.049	0.046
50	0.90	0.061	0.038	0.068	0.048
100	0.90	0.143	0.129	0.202	0.207
150	0.90	0.370	0.318	0.430	0.391
300	0.90	0.942	0.863	0.968	0.932
50	0.85	0.064	0.044	0.080	0.078
100	0.85	0.225	0.227	0.358	0.353
150	0.85	0.612	0.518	0.711	0.638
300	0.85	0.995	0.974	0.999	0.995
50	0.80	0.067	0.061	0.096	0.102
100	0.80	0.328	0.331	0.515	0.498
150	0.80	0.789	0.684	0.880	0.820
300	0.80	0.999	0.994	1.000	0.999

Note:

1. DGP is as described in (1) and (2) where we set $\beta_0 = \beta_1 = 0$. $v_t = \phi v_{t-1} + \epsilon_{t-1}$, with $\phi = 0.5$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, and $\alpha = 1, 0.9, 0.85, 0.8$.
2. The replications for the asymptotic tests are 5000, while for the bootstrap tests are 1000 (=NB).
3. The 5%-level asymptotic critical value of the $MZ - GLS$ test is -8.1 for $m = 0$ and -17.3 for $m = 1$.
4. The lag orders p fit to the bootstrap ADF regression and to the bootstrap autoregression in reproducing the samples are separately chosen by both AIC and BIC. The maximum lag orders are set to be 5.

Table 6: Size and Power of Unit Root Statistics, MZ

Panel A: $m=0$ and AR(1) error

T	α	MZ	MZ^*	MZ	MZ^*
		AIC	AIC	BIC	BIC
50	1	0.118	0.033	0.072	0.040
100	1	0.068	0.041	0.047	0.041
150	1	0.054	0.047	0.041	0.051
300	1	0.043	0.037	0.036	0.038
50	0.90	0.106	0.102	0.155	0.145
100	0.90	0.344	0.297	0.419	0.399
150	0.90	0.629	0.547	0.708	0.636
300	0.90	0.988	0.966	0.997	0.988
50	0.85	0.127	0.157	0.221	0.213
100	0.85	0.525	0.442	0.637	0.589
150	0.85	0.856	0.745	0.918	0.869
300	0.85	0.999	0.997	1.000	1.000
50	0.80	0.147	0.188	0.283	0.268
100	0.80	0.678	0.584	0.796	0.728
150	0.80	0.945	0.861	0.981	0.954
300	0.80	1.000	1.000	1.000	1.000

Panel B: $m=1$ and AR(1) error

T	α	MZ	MZ^*	MZ	MZ^*
		AIC	AIC	BIC	BIC
50	1	0.248	0.027	0.129	0.033
100	1	0.129	0.037	0.068	0.043
150	1	0.092	0.041	0.057	0.059
300	1	0.060	0.037	0.045	0.044
50	0.90	0.056	0.036	0.064	0.049
100	0.90	0.120	0.108	0.176	0.179
150	0.90	0.275	0.261	0.370	0.331
300	0.90	0.878	0.791	0.931	0.885
50	0.85	0.058	0.047	0.072	0.068
100	0.85	0.170	0.183	0.302	0.305
150	0.85	0.472	0.446	0.626	0.572
300	0.85	0.985	0.938	0.997	0.986
50	0.80	0.060	0.062	0.080	0.101
100	0.80	0.230	0.279	0.436	0.440
150	0.80	0.657	0.590	0.819	0.750
300	0.80	0.997	0.985	1.000	0.998

Note:

1. DGP is as described in (1) and (2) where we set $\beta_0=\beta_1=0$. $v_t=\phi v_{t-1}+\epsilon_{t-1}$, with $\phi=0.5$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, and $\alpha=1, 0.9, 0.85, 0.8$.
2. The replications for the asymptotic tests are 5000, while for the bootstrap tests are 1000 (=NB).
3. The 5%-level asymptotic critical value of the MZ test is -14.1 for $m=0$ and -21.3 for $m=1$.
4. The lag orders p fit to the bootstrap ADF regression and to the bootstrap autoregression in reproducing the samples are separately chosen by both AIC and BIC. The maximum lag orders are set to be 5.

Table 7: Size and Power of Unit Root Statistics, $MZ-GLS$

Panel A: $m=0$ and MA(1) error

T	α	$MZ-GLS$	$MZ-GLS^*$	$MZ-GLS$	$MZ-G$
		AIC	AIC	BIC	BIC
50	1	0.188	0.060	0.220	0.09
100	1	0.131	0.066	0.183	0.11
150	1	0.103	0.078	0.160	0.11
300	1	0.078	0.053	0.120	0.08
50	0.90	0.169	0.179	0.185	0.32
100	0.90	0.379	0.447	0.424	0.60
150	0.90	0.559	0.659	0.618	0.75
300	0.90	0.851	0.864	0.873	0.90
50	0.85	0.242	0.231	0.274	0.41
100	0.85	0.502	0.554	0.576	0.69
150	0.85	0.648	0.727	0.723	0.81
300	0.85	0.851	0.878	0.880	0.90
50	0.80	0.312	0.296	0.362	0.46
100	0.80	0.559	0.606	0.657	0.72
150	0.80	0.664	0.738	0.755	0.83
300	0.80	0.832	0.873	0.867	0.89

Panel B: $m=1$ and MA(1) error

T	α	$MZ-GLS$	$MZ-GLS^*$	$MZ-GLS$	$MZ-G$
		AIC	AIC	BIC	BIC
50	1	0.182	0.042	0.162	0.09
100	1	0.145	0.060	0.200	0.14
150	1	0.117	0.069	0.199	0.13
300	1	0.072	0.069	0.141	0.11
50	0.90	0.072	0.065	0.095	0.15
100	0.90	0.209	0.223	0.235	0.39
150	0.90	0.373	0.429	0.424	0.60
300	0.90	0.821	0.874	0.845	0.91
50	0.85	0.090	0.080	0.134	0.20
100	0.85	0.349	0.366	0.413	0.55
150	0.85	0.558	0.633	0.663	0.76
300	0.85	0.890	0.930	0.918	0.95
50	0.80	0.108	0.114	0.184	0.26
100	0.80	0.480	0.460	0.583	0.63
150	0.80	0.662	0.750	0.781	0.84
300	0.80	0.895	0.942	0.931	0.96

Note:

1. DGP is as described in (1) and (2) where we set $\beta_0=\beta_1=0$. $v_t=\epsilon_t+\theta\epsilon_{t-1}$, with $\theta=-0.5$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, and $\alpha=1, 0.9, 0.85, 0.8$.
2. The replications for the asymptotic tests are 5000, while for the bootstrap tests are 1000 (=NB).
3. The 5%-level asymptotic critical value of the $MZ-GLS$ test is -8.1 for $m=0$ and -17.3 for $m=1$.
4. The lag orders p fit to the bootstrap ADF regression and to the bootstrap autoregression in reproducing the samples are separately chosen by both AIC and BIC. The maximum lag orders are set to be 5.

Table 8: Size and Power of Unit Root Statistics, MZ

Panel A: $m = 0$ and MA(1) error

T	α	MZ AIC	MZ^* AIC	MZ BIC	MZ^* BIC
50	1	0.110	0.052	0.107	0.101
100	1	0.101	0.075	0.142	0.119
150	1	0.084	0.065	0.140	0.131
300	1	0.057	0.062	0.106	0.090
50	0.90	0.153	0.150	0.183	0.307
100	0.90	0.417	0.413	0.460	0.629
150	0.90	0.623	0.718	0.705	0.844
300	0.90	0.985	0.991	0.988	0.997
50	0.85	0.232	0.211	0.283	0.418
100	0.85	0.637	0.631	0.718	0.799
150	0.85	0.832	0.909	0.894	0.942
300	0.85	0.999	0.999	0.999	1.000
50	0.80	0.322	0.280	0.400	0.493
100	0.80	0.781	0.769	0.864	0.869
150	0.80	0.921	0.975	0.965	0.985
300	0.80	1.000	1.000	1.000	1.000

Panel B: $m = 1$ and MA(1) error

T	α	MZ AIC	MZ^* AIC	MZ BIC	MZ^* BIC
50	1	0.165	0.039	0.110	0.093
100	1	0.146	0.052	0.175	0.149
150	1	0.124	0.062	0.199	0.135
300	1	0.082	0.076	0.158	0.122
50	0.90	0.067	0.057	0.093	0.148
100	0.90	0.204	0.213	0.229	0.392
150	0.90	0.384	0.391	0.436	0.605
300	0.90	0.870	0.927	0.876	0.968
50	0.85	0.078	0.083	0.129	0.192
100	0.85	0.364	0.348	0.421	0.569
150	0.85	0.640	0.641	0.736	0.791
300	0.85	0.975	0.994	0.986	0.995
50	0.80	0.090	0.108	0.181	0.245
100	0.80	0.544	0.481	0.637	0.685
150	0.80	0.797	0.833	0.888	0.902
300	0.80	0.994	1.000	0.998	0.999

Note:

1. DGP is as described in (1) and (2) where we set $\beta_0 = \beta_1 = 0$. $v_t = \epsilon_t + \theta\epsilon_{t-1}$, with $\theta = 0.5$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, and $\alpha = 1, 0.9, 0.85, 0.8$.
2. The replications for the asymptotic tests are 5000, while for the bootstrap tests are 1000 (=NB).
3. The 5%-level asymptotic critical value of the MZ test is -14.1 for $m = 0$ and -21.3 for $m = 1$.
4. The lag orders p fit to the bootstrap ADF regression and to the bootstrap autoregression in reproducing the samples are separately chosen by both AIC and BIC. The maximum lag orders are set to be 5.