Groupe de Recherche en Économie et Développement International

Cahier de Recherche / Working Paper
07-05

ARMA Sieve bootstrap unit root tests

Patrick Richard

UNIVERSITÉ DE SHERBROOKE

ARMA Sieve Bootstrap Unit Root Tests

Patrick Richard*

Université de Sherbrooke and GREDI
patrick.richard2@usherbrooke.ca
(819) 821-8000 ext:63233

July 2009

Abstract

Augmented Dickey-Fuller unit root tests may severely overreject when the DGP is a general linear process. The use of the AR sieve bootstrap, proposed by Park (2002) and Chang and Park (2003), may alleviate this problem. We propose sieve bootstraps based on MA and ARMA approximations. Invariance principles for the partial sum processes built from these sieve bootstrap DGPs are established and a proof of the asymptotic validity of the resulting ADF bootstrap tests is provided. Through Monte Carlo experiments, we find that the rejection probabilities of the MA and ARMA sieve bootstraps are often lower and more robust to the underlying DGP than that of the AR sieve bootstrap. In particular, the new sieve bootstraps perform much better than the AR sieve when a large MA root is present. We also find that the ARMA sieve bootstrap requires only a very parsimonious specification to achieve excellent results.

* I thank Russell Davidson, John Galbraith, Nikolay Gospodinov, Alex Maynard, James MacKinnon, Christos Ntantamis, David Stephens, Victoria Zinde-Walsh and seminar participants at the Econometric Society NASM 2007 at Duke University, CEF 2007 at HEC Montréal, SCSE 2007 in Québec city and CEA 2006 at Concordia University conferences for helpful comments on various versions of this paper.

1 Introduction

It is well known that the Augmented Dickey-Fuller (ADF) test sometimes severely overrejects the unit root null hypothesis (see, for example, Schwert, 1989). The bootstrap is regularly used as a potential solution to such error in rejection probability (ERP) problems because it often provides asymptotic refinements over standard asymptotic tests. Park (2003) shows that such refinements exist for bootstrap ADF tests conducted on unit root time series. However, his results only apply to DGPs where the first difference process is a finite order AR(p) with p known, so that the bootstrap problem may easily be reduced to *iid* resampling from the residuals' empirical density function (EDF). They consequently do not apply to more general time series process of possibly infinite order and unknown form.

Nevertheless, ADF tests based on autoregressive sieve bootstraps have been shown to be asymptotically valid (Psaradakis, 2001, Park, 2002 and Chang and Park, 2003), meaning that the tests statistics asymptotic distribution is the standard Dickey-Fuller. Furthermore, the simulation experiments carried out by these authors indicate that such sieve bootstrap tests may have much smaller ERP than tests based on analytical critical values.

This paper proposes sieve bootstraps based on MA and ARMA approximations. To this date the sieve bootstrap, originally introduced by Kreiss (1992) and Bühlmann (1997, 1998), has always relied on AR approximations. We show, using a methodology very closely related to that of Park (2002) and Chang and Park (2003), that ADF tests based on MA and ARMA sieve bootstraps are asymptotically valid as well. Our simulations show that MA and ARMA sieve bootstrap tests may sometimes be more accurate that AR sieve bootstrap tests.

The rest of the paper is organised as follows. In section 2, we provide invariance principles for partial sum processes built with MA and ARMA sieve bootstrap data. This is then used to show that ADF tests based on either of these sieves are asymptotically valid. Section 3 studies the finite sample performances of these tests and compares them to the more usual AR sieve bootstrap ADF tests. Section 4 concludes.

2 Invariance Principle and Validity of Sieve Bootstrap ADF Tests

In this section, we derive the invariance principles necessary to justify the use of MA sieve bootstrap (MASB) and ARMA sieve bootstrap (ARMASB) methods to carry out bootstrap unit root tests. We also show that ADF tests based on the MASB and ARMASB are asymptotically valid, that is, that their asymptotic distribution is the standard Dickey-Fuller. Although these results are derived for unit root models without a constant or a deterministic trend, the proofs could easily be modified to allow for such deterministic parts.

Establishing invariance principles for sieve bootstrap procedures is a relatively new strand in the literature and, at the time of writing this, and to the author's knowledge, only two attempts have been made. The first is Bickel and Bühlmann (1999) who derive a bootstrap functional central limit theorem for the AR sieve bootstrap (ARSB). The second is Park (2002) who derives an invariance principle for the ARSB with data generated by a general MA(∞) linear process. Most of what we do here borrows heavily from this second approach.

2.1 General bootstrap invariance principle

Let $\{\varepsilon_t\}_{t=1}^n$ be a sequence of *iid* random variables with finite variance σ^2. Consider a sample of size n and define the partial sum process:

$$W_n(t) = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{[nt]} \varepsilon_k$$

where $[y]$ denotes the largest integer smaller or equal to y and t is an index such that $\frac{j-1}{n} \leq t < \frac{j}{n}$, where $j = 1, 2, ...n$ is another index that allows us to divide the [0,1] interval into $n+1$ parts. Thus, $W_n(t)$ is a step function that converges to a random walk as $n \to \infty$. Also, as $n \to \infty$, W_n(t) becomes infinitely dense on the $[0, 1]$ interval. By the classical Donsker's theorem, we know that

$$W_n \xrightarrow{d} W.$$

where W is the standard Brownian motion. The Skorohod representation theorem tells us that there exists a probability space (Ω,$\mathcal{F}$,P) that supports W and a process W_n' such that W_n' has the same distribution as W_n and

$$W_n' \xrightarrow{a.s.} W. \tag{1}$$

Furthermore, as demonstrated by Sakhanenko (1980), W_n' can be chosen so that

$$P\{\|W_n' - W\| \geq \delta\} \leq n^{1-r/2} K_r E\left(|\varepsilon_t|^r\right) \tag{2}$$

for any $\delta > 0$ and $r > 2$ such that $E(|\varepsilon_t|^r) < \infty$ and where K_r is a constant that depends on r only. Because the invariance principle we seek to establish is a distributional result, we do not need to distinguish W_n from W_n'. Consequently, because of (1), we say that $W_n \xrightarrow{a.s.} W$, which is stronger than the convergence in distribution implied by Donsker's theorem.

Now, suppose that we can obtain an estimate of $\{\varepsilon_t\}_{t=1}^n$, which we will denote as $\{\hat{\varepsilon}_t\}_{t=1}^n$, from which we can draw bootstrap samples of size n, denoted as $\{\varepsilon_t^\star\}_{t=1}^n$. If we suppose that $n \to \infty$, then we can build a bootstrap probability space $(\Omega^\star, \mathcal{F}^\star, P^\star)$ which is conditional on the realization of the set of residuals $\{\hat{\varepsilon}_t\}_{t=1}^\infty$ from which the bootstrap random variables are drawn. What this means is that each bootstrap drawing $\{\varepsilon_t^\star\}_{t=1}^n$ can be seen as a realization of a random variable defined on $(\Omega^\star, \mathcal{F}^\star, P^\star)$.

In all that follows, the expectation with respect to this space (that is, with respect to the probability $P^\star$) will be denoted by $E^\star$. For example, if the bootstrap samples are drawn from $\{(\hat{\varepsilon}_t - \bar{\varepsilon}_n)\}_{t=1}^n$, then $E^\star(\varepsilon_t^\star) = 0$ and $E^\star\left(\varepsilon_t^{\star 2}\right) = \hat{\sigma}_n^2 = (1/n)\sum_{t=1}^n (\hat{\varepsilon}_t - \bar{\varepsilon}_n)^2$. Of course, whenever the $\hat{\varepsilon}_t$s are residuals from a linear regression model with a constant, $\bar{\varepsilon}_n = 0$ so that $E^\star\left(\varepsilon_t^{\star 2}\right) = \hat{\sigma}_n^2 = (1/n)\sum_{t=1}^n \hat{\varepsilon}_t^2$. Also, $\stackrel{d^\star}{\to}$, $\stackrel{p^\star}{\to}$ and $\stackrel{a.s.^\star}{\to}$ will be used to denote convergence in distribution, in probability and almost sure of the functionals of the bootstrap samples defined on $(\Omega^\star, \mathcal{F}^\star, P^\star)$. Further, following Park (2002), for any sequence of bootstrapped statistics $\{X_n^\star\}$ we say that $X_n^\star \stackrel{d^\star}{\to} X$ a.s. if the conditional distribution of $\{X_n^\star\}$ weakly converges to that of X a.s on all sets of $\{\hat{\varepsilon}_t\}_{t=1}^\infty$. In other words, if the bootstrap convergence in distribution ($\stackrel{d^\star}{\to}$) of functionals of bootstrap samples on $(\Omega^\star, \mathcal{F}^\star, P^\star)$ happens almost surely for all realizations of $\{\hat{\varepsilon}_t\}_{t=1}^\infty$, then we write $\stackrel{d^\star}{\to}$ a.s.

Let $\{\varepsilon_t^\star\}_{t=1}^n$ be a realization from a bootstrap probability space. Define

$$W_n^\star(t) = \frac{1}{\hat{\sigma}_n \sqrt{n}} \sum_{k=1}^{[nt]} \varepsilon_k^\star.$$

Once again, by Skorohod and Sakhanenko's theorems, there exists a probability space on which a Brownian motion $W^\star$ is supported and on which there also exists a process $W_n^{\star\prime}$ which has the same distribution as $W_n^\star$ and such that

$$P^\star\{\|W_n^{\star\prime} - W^\star\| \geq \delta\} \leq n^{1-r/2} K_r E^\star\left(|\varepsilon_t^\star|^r\right) \tag{3}$$

for δ, r and K_r defined as before. Because $W_n^{\star\prime}$ and $W_n^\star$ are distributionally equivalent, we will not distinguish them in all that follows. The inequality (3) allows us to state the following theorem, which is also theorem 2.2 in Park (2002):

Theorem (Park 2002, theorem 2.2, p. 473). *If $E^\star\left(|\varepsilon_t^\star|^r\right) < \infty$ a.s. and*

$$n^{1-r/2} E^\star\left(|\varepsilon_t^\star|^r\right) \stackrel{a.s.}{\to} 0 \tag{4}$$

for some $r > 2$, then $W_n^\star \stackrel{d^\star}{\to} W$ a.s. as $n \to \infty$.

This result comes from the fact that, if condition (4) holds, then inequality (3) implies convergence in probability over the bootstrap probability space which, as usual, implies convergence in distribution that is, $W_n^\star \xrightarrow{d^\star} W^\star$ a.s. Since the distribution of $W^\star$ is independent of the set of residuals $\{\hat{\varepsilon}_t\}_{t=1}^{\infty}$, we can equivalently say $W_n^\star \xrightarrow{d^\star} W$ a.s. Hence, whenever condition (4) is met, the invariance principle follows. Of course, this theorem is only valid for bootstrap samples drawn from sets of *iid* random variables. Nevertheless, Park (2002) uses it to prove an invariance principle for the AR sieve bootstrap. In the next two subsections, we do essentially the same thing for the MA and ARMA sieve bootstraps.

2.2 Invariance principle for MA sieve bootstrap

Let us consider a general linear process:

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

where

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

where $\pi(z)$ and ε_t satisfy the following assumptions:

Assumption 1.

(a) The ε_ts are iid random variables such that $E(\varepsilon_t) = 0$ and $E(|\varepsilon_t|^r) < \infty$ for some $r > 4$.

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

See Park (2002) and Chang and Park (2003) for a discussion of these assumptions. The MASB consists of approximating equation (5) by a finite order MA(q) model:

$$u_t = \pi_1 \varepsilon_{q,t-1} + \pi_2 \varepsilon_{q,t-2} + \ldots + \pi_q \varepsilon_{q,t-q} + \varepsilon_{q,t} \tag{6}$$

where q is a function of the sample size. Approximating an infinite linear process by a finite dimension model is an old topic in econometrics. Most of the time, finite f-order autoregressions are used, with f increasing as a function of the sample size. The classical reference on the subject is Berk (1974). Galbraith and Zinde-Walsh (1994, 1997), henceforth GZW(1994) and GZW(1997) use analytical binding functions between the parameters of this AR(f) approximation and the parameters of any finite order MA(q) or ARMA(p, q) process to estimate the parameters of these latter models. Some of our proofs make use of these binding functions.

Assumption 2.

$q \to \infty$ and $f \to \infty$ as $n \to \infty$ and $q = o\left((n/\log(n))^{1/2}\right)$ and $f = o\left((n/\log(n))^{1/2}\right)$ with $f > q$.

The reason for this choice is closely related to lemma 3.1 in Park (2002) and the reader is referred to the discussion following it. Here, we limit ourselves to pointing out that this rate is consistent with both AIC and BIC, which are commonly used in practice. The restriction that $f > q$ is necessary for the computation of the GZW (1994) and GZW (1997) estimators.

The bootstrap samples are generated from the DGP:

$$u_t^\star = \hat{\pi}_{q,1}\varepsilon_{t-1}^\star + ... + \hat{\pi}_{q,q}\varepsilon_{t-q}^\star + \varepsilon_t^\star. \tag{7}$$

where the $\hat{\pi}_{q,i}$, $i = 1, 2, ...q$ are estimators of the true parameters π_i and the $\varepsilon_t^\star$ are drawn from the EDF of $(\hat{\varepsilon}_t - \bar{\varepsilon}_n)$, that is, from the EDF of the centered residuals of the MA(q) sieve. We will now establish an invariance principle for the partial sum process of $u_t^\star$ by considering its Beveridge-Nelson decomposition and showing that it converges almost surely to the same limit as the corresponding partial sum process built with the original u_t. First, consider the decomposition of u_t:

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

where

$$\tilde{u}_t = \sum_{k=0}^{\infty} \tilde{\pi}_k \varepsilon_{t-k}$$

and

$$\tilde{\pi}_k = \sum_{i=k+1}^{\infty} \pi_i.$$

Now, consider the partial sum process

$$V_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} u_k$$

$$= \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} \pi(1)\varepsilon_t + \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} \sum_{k=0}^{\infty} \left[\left(\sum_{i=k+1}^{\infty} \pi_i \right) (\varepsilon_{t-k-1} - \varepsilon_{t-k}) \right]$$

hence,

$$V_n(t) = (\sigma\pi(1))W_n(t) + \frac{1}{\sqrt{n}}(\tilde{u}_0 - \tilde{u}_{[nt]}).$$

Under assumption 1, Phillips and Solo (1992) show that

$$\max_{1 \le k \le n} |n^{-1/2}\tilde{u}_k| \xrightarrow{p} 0.$$

Therefore, applying the continuous mapping theorem, we have

$$V_n(t) \xrightarrow{d} V = (\sigma\pi(1))W.$$

On the other hand, from equation (7), we see that $u_t^\star$ can be decomposed as

$$u_t^\star = \hat{\pi}(1)\varepsilon_t^\star + (\tilde{u}_{t-1}^\star - \tilde{u}_t^\star)$$

where

$$\hat{\pi}(1) = 1 + \sum_{k=1}^{q} \hat{\pi}_{q,k}$$

$$\tilde{u}_t^\star = \sum_{k=1}^{q} \tilde{\hat{\pi}}_k \varepsilon_{t-k+1}^\star$$

$$\tilde{\hat{\pi}}_k = \sum_{i=k}^{q} \hat{\pi}_{q,i}.$$

It therefore follows that we can write:

$$V_n^\star(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} u_t^\star = (\hat{\sigma}_n \hat{\pi}_n(1))W_n^\star + \frac{1}{\sqrt{n}}(\tilde{u}_0^\star - \tilde{u}_{[nt]}^\star).$$

In order to establish the invariance principle, we must show that $V_n^\star \xrightarrow{d^\star} V = (\sigma\pi(1))W$ a.s. To do this, we need 3 lemmas. The first one shows that $\hat{\sigma}_n$ and $\hat{\pi}_n(1)$ converge almost surely to σ and $\pi(1)$. The second demonstrates that $W_n^\star(t) \xrightarrow{d^\star} W$ a.s. Finally, the last one shows that

$$P^\star\{\max_{1\leq t\leq n} |n^{-1/2}\tilde{u}_t^\star| > \delta\} \xrightarrow{a.s.} 0 \tag{8}$$

for all $\delta > 0$, which is equivalent to saying that

$$\max_{1\leq k\leq n} |n^{-1/2}\tilde{u}_k| \xrightarrow{p^\star} 0 \text{ a.s.}$$

and is therefore the bootstrap equivalent of the result of Philips and Solo (1992). These 3 lemmas are closely related to the results of Park (2002) and their counterparts in that paper are identified for reference. The proofs can be found in the appendix.

Lemma 1 (Park 2002, lemma 3.1, p. 476). *Let assumptions 1 and 2 hold. Then,*

$$\max_{1\leq k\leq q} |\hat{\pi}_{q,k} - \pi_k| = o(1) \quad \text{a.s.} \tag{9}$$

$$\hat{\sigma}_n^2 = \sigma^2 + o(1) \quad \text{a.s.} \tag{10}$$

$$\hat{\pi}_n(1) = \pi(1) + o(1) \quad \text{a.s.} \tag{11}$$

Lemma 2 (Park 2002, lemma 3.2, p. 477). *Let assumptions 1 and 2 hold. Then,*

$$E^\star(|\varepsilon_t^\star|^r) < \infty \text{ a.s. and } n^{1-r/2}E^\star(|\varepsilon_t^\star|^r) \overset{a.s.}{\to} 0$$

Lemma 2 proves that $W_n^\star(t) \overset{d^\star}{\to} W$ a.s. because it shows that condition (4) holds almost surely.

Lemma 3 (Park 2002, theorem 3.3, p. 478). *Let assumptions 1 and 2 hold. Then, (8) holds.*

With these 3 lemmas, the MASB invariance principle is established. It is formalized in the next theorem:

Theorem 1. *Let assumptions 1 and 2 hold. Then by lemmas 1, 2 and 3,*

$$V_n^\star \overset{d^\star}{\to} V = (\sigma\pi(1))W \quad \text{a.s.}$$

2.3 Invariance principle for ARMA sieve bootstrap

Establishing an invariance principle for the ARMASB is a simple matter of combining the results of the previous subsection to those of Park (2002). The ARMASB procedure consists of approximating the general linear process (5) by a finite order ARMA(p,q) model:

$$u_t = \alpha_1 u_{t-1} + ... + \alpha_p u_{t-p} + \pi_1\varepsilon_{\ell,t-1} + ... + \pi_q\varepsilon_{\ell,t-q} + \varepsilon_{\ell,t} \tag{12}$$

where $\ell = p + q$ denotes the total number of parameters and is, of course, a function of the sample size. Our proofs use the binding functions of GZW (1997). Hence, in addition to p and q, we must also specify the order of an approximating autoregression. As before, we let f denote this order. Then, we make the following assumptions:

Assumption 3

Both ℓ and f go to infinity at the rate $o\left((n/\log(n))^{1/2}\right)$ and $f > \ell$.

Notice that we do not require that both p and q go to infinity simultaneously. Rather, we require that their sum does. Thus, the results that follow hold even if p or q is held fixed while the sample size increases, as long as the sum increases at the proper rate. The bootstrap samples are generated from the DGP:

$$u_t^\star = \hat{\alpha}_{\ell,1}u_{t-1}^\star + ... + \hat{\alpha}_{\ell,p}u_{t-p}^\star + \hat{\pi}_{\ell,1}\varepsilon_{t-1}^\star + ... + \hat{\pi}_{\ell,q}\varepsilon_{t-q}^\star + \varepsilon_t^\star \tag{13}$$

where the $\hat{\alpha}_{\ell,k}$ and $\hat{\pi}_{\ell,k}$ are consistent estimators and the $\varepsilon_t^\star$ are drawn from the EDF of $(\hat{\varepsilon}_t - \bar{\varepsilon}_n)_{t=1}^n$, that is, from the empirical distribution of the centered residuals of the fitted ARMA(p,q). Evidently, $\hat{\pi}_{\ell,i}$ and $\hat{\alpha}_{\ell,i}$ would not be the same for every combinations of p and q such that $p+q=\ell$ as our notation would suggest. We nevertheless keep this slight inaccuracy for the sake of simplicity. Next, we need to build a partial sum process $V_n^\star$ of $u_t^\star$. The easiest way to do this is to consider either the AR(∞) or the MA(∞) form of the ARMA(p,q) model and build $V_n^\star$ based on this representation. Let us consider the MA(∞) form of $u_t^\star$, which we define as

$$u_t^\star = \hat{\theta}_{\ell,1}\varepsilon_{t-1}^\star + \hat{\theta}_{\ell,2}\varepsilon_{t-2}^\star + ... + \varepsilon_t^\star \tag{14}$$

where $\hat{\theta}_{\ell,1} = \hat{\pi}_{\ell,1} + \hat{\alpha}_{\ell,1}$, $\hat{\theta}_{\ell,2} = \hat{\theta}_{\ell,1}\hat{\alpha}_{\ell,1} - \hat{\alpha}_{\ell,2} - \hat{\pi}_{\ell,2}$ and so forth. Then,

$$V_n^\star(t) = \frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]} u_t^\star = (\hat{\sigma}_n\hat{\theta}_n(1))W_n^\star + \frac{1}{\sqrt{n}}(\tilde{u}_0^\star - \tilde{u}_{[nt]}^\star) \tag{15}$$

where

$$\tilde{u}_t^\star = \sum_{k=0}^{\infty}\tilde{\theta}_k\varepsilon_{t-k}^\star$$

and

$$\tilde{\theta}_k = \sum_{i=k+1}^{\infty}\hat{\theta}_i.$$

and where $\hat{\sigma}_n$ is the estimated variance of the residuals of the ARMA(p,q) sieve. Then, we need to show that $V_n^\star(t) \xrightarrow{d^\star} V$ a.s. This, as before, can be done through proving 3 results, which are simple corollaries of lemmas 1, 2 and 3 of the present paper and lemmas 3.1 and 3.2 as well as theorem 3.3 of Park (2002).

Corollary 1. *Under assumptions 1 and 3,*

$$\max_{1\le k\le \ell}\left|\hat{\theta}_{\ell,k} - \pi_k\right| = o(1) \quad \text{a.s.} \tag{16}$$

$$\hat{\sigma}_n^2 = \sigma^2 + o(1) \quad \text{a.s.} \tag{17}$$

$$\hat{\theta}_n(1) = \pi(1) + o(1) \quad \text{a.s.} \tag{18}$$

Corollary 2. *Under assumptions 1 and 3,*

$$W_n^\star \xrightarrow{d^\star} W \text{ a.s.}$$

Corollary 3. *Under assumptions 1 and 3,*

$$P^\star\{\max_{1\le t\le n}|n^{-1/2}\tilde{u}_t^\star| > \delta\} \xrightarrow{a.s.} 0$$

for all $\delta > 0$ and $u_t^\star$ generated from the ARMASB DGP.

These three results, the proofs of which may be found in the appendix, are sufficient to prove the invariance principle of the ARMASB partial sum process.

Theorem 2. *Let assumptions 1, 2 and 3 hold. Then by corollaries 1, 2 and 3,*

$$V_n^\star \xrightarrow{d^\star} V = (\sigma\pi(1))W \quad \text{a.s.}$$

2.4 Asymptotic Validity of MASB and ARMASB ADF tests

Consider a time series y_t with the following DGP:

$$y_t = \alpha y_{t-1} + u_t \tag{19}$$

where u_t is the general linear process described in equation (5). We want to test the unit root hypothesis against the stationarity alternative (that is, $\mathrm{H}_0 : \alpha = 1$ against $\mathrm{H}_1 : \alpha < 1$). This test is frequently conducted as a t-test in the so called ADF regression, first proposed by Dickey and Fuller (1979, 1981):

$$y_t = \alpha y_{t-1} + \sum_{k=1}^{p} \alpha_k \Delta y_{t-k} + e_{p,t} \tag{20}$$

where p is chosen as a function of the sample size. Deterministic parts such as a constant and a time trend are usually added to the regressors of (20). Chang and Park (2002) have shown that the test based on this regression asymptotically follows the DF distribution when H_0 is true under very weak conditions, including assumptions 1 and 2. Let $y_t^\star$ denote the bootstrap process such that:

$$y_t^\star = \sum_{k=1}^{t} u_k^\star$$

where the $u_t^\star = \Delta y_t^\star$ are generated by some sieve bootstrap DGP. The bootstrap ADF regression equivalent to regression (20) is

$$y_t^\star = \alpha y_{t-1}^\star + \sum_{k=1}^{p} \alpha_k \Delta y_{t-k}^\star + e_t. \tag{21}$$

Let us suppose for a moment that $\Delta y_t^\star$ has been generated by an AR(p) sieve bootstrap DGP:

$$\Delta y_t^\star = \sum_{k=1}^{p} \hat\alpha_{p,k} \Delta y_{t-k}^\star + \varepsilon_t^\star$$

Then, letting $\alpha = 1$ in (21), we see that the true parameters of this equation are the $\hat\alpha_{p,k}$s and that its errors are identical to the errors driving the bootstrap DGP. This is a convenient fact which Chang and Park (2003, here after CP (2003)) use to prove the consistency of the ARSB ADF test based on this regression. If however the $y_t^\star$ are generated by the MASB or ARMASB described above, then the errors of regression (21) are not identical to the bootstrap errors under the null. It is nevertheless possible to show that they will be equivalent asymptotically, that is, that $\varepsilon_t^\star = e_t + o(1)$ a.s. This is done in lemma A1, which can be found in the appendix.

Let $x_{p,t}^\star = (\Delta y_{t-1}^\star, \Delta y_{t-2}^\star, ..., \Delta y_{t-p}^\star)$ and define:

$$G_n^\star = \sum_{t=1}^{n} y_{t-1}^\star \varepsilon_t^\star - \left(\sum_{t=1}^{n} y_{t-1}^\star {x_{p,t}^\star}^\top\right)\left(\sum_{t=1}^{n} x_{p,t}^\star {x_{p,t}^\star}^\top\right)^{-1}\left(\sum_{t=1}^{n} x_{p,t}^\star \varepsilon_t^\star\right)$$

$$H_n^\star = \sum_{t=1}^{n} {y_{t-1}^\star}^2 - \left(\sum_{t=1}^{n} y_{t-1}^\star {x_{p,t}^\star}^\top\right)\left(\sum_{t=1}^{n} x_{p,t}^\star {x_{p,t}^\star}^\top\right)^{-1}\left(\sum_{t=1}^{n} x_{p,t}^\star y_{t-1}^\star\right)$$

which are products of orthogonal projections of $y_{t-1}^\star$ and $\varepsilon_t^\star$ on to the orthogonal complement of the space generated by $x_{p,t}^\star$. Then, the t-statistic computed from regression (21) can be written as:

$$\tau_n^\star = \frac{\hat{\alpha}_n^\star - 1}{s(\hat{\alpha}_n^\star)} + o(1) \text{ a.s.}$$

where $\hat{\alpha}_n^\star - 1 = G_n^\star H_n^{\star -1}$ and $s(\hat{\alpha}_n^\star)^2 = \hat{\sigma}_n^2 H_n^{\star -1}$. The equality is asymptotic and almost surely holds because of the results of lemma A1. This lemma also justifies the use of the estimated variance $\hat{\sigma}_n^2$. Note that in small samples, it may be preferable to use the estimated variance of the residuals from the ADF regression, which is indeed what we do in the simulations presented below.

Assumption 4.

$q = c_q n^k$, $\ell = c_\ell n^k$ $p = c_p n^k$ where c_q, c_ℓ and c_p are constants and $1/rs < k < 1/2$ and p is the order of the ADF regression.

Assumptions 2 and 3 can be fitted into this assumption for appropriate values of k. Also, notice that assumption 4 imposes a lower bound on the growth rate of p, ℓ and q. This is necessary to obtain almost sure convergence. See CP (2003) for a weaker assumption that allows for convergence in probability. Several preliminary and quite technical results are necessary to prove that the bootstrap test based on the statistic $\tau_n^\star$ is consistent. To avoid rendering the present exposition more laborious than it needs to be, we relegate them to the appendix (lemmas A2 to A5).

2.4.1 Asymptotic validity of MASB ADF test

In order to show that the MASB ADF test is consistent, we now prove two results on the elements of $G_n^\star$ and $H_n^\star$. Those results are stated in terms of bootstrap stochastic orders, denoted by $O_p^\star$ and $o_p^\star$, see CP (2003) for details.

Lemma 4. *Under assumptions 1 and 4, we have*

$$\frac{1}{n}\sum_{t=1}^{n} y_{t-1}^\star \varepsilon_t^\star = \hat{\pi}_n(1)\frac{1}{n}\sum_{t=1}^{n} w_{t-1}^\star \varepsilon_t^\star + o_p^\star(1) \text{ a.s.} \tag{22}$$

$$\frac{1}{n^2}\sum_{t=1}^{n} {y_{t-1}^\star}^2 = \hat{\pi}_n(1)^2\frac{1}{n^2}\sum_{t=1}^{n} {w_{t-1}^\star}^2 + o_p^\star(1) \text{ a.s.} \tag{23}$$

where $w_t^\star = \sum_{k=1}^{t} \varepsilon_k^\star$.

Lemma 5. *Under assumptions 1 and 4 we have*

$$\left\| \left(\frac{1}{n}\sum_{t=1}^{n} x_{p,t}^\star {x_{p,t}^\star}^\top \right)^{-1} \right\| = O_p^\star(1) \text{ a.s.} \tag{24}$$

$$\left\| \sum_{t=1}^{n} x_{p,t}^\star y_{t-1}^\star \right\| = O_p^\star(np^{1/2}) \text{ a.s.} \tag{25}$$

$$\left\| \sum_{t=1}^{n} x_{p,t}^\star \varepsilon_t^\star \right\| = O_p^\star(n^{1/2}p^{1/2}) \text{ a.s.} \tag{26}$$

We can place an upper bound on the absolute value of the second term of $G_n^\star$:

$$\left| \left(\sum_{t=1}^{n} y_{t-1}^\star {x_{p,t}^\star}^\top \right) \left(\sum_{t=1}^{n} x_{p,t}^\star {x_{p,t}^\star}^\top \right)^{-1} \left(\sum_{t=1}^{n} x_{p,t}^\star \varepsilon_t^\star \right) \right| \le \left\| \sum_{t=1}^{n} y_{t-1}^\star {x_{p,t}^\star}^\top \right\| \left\| \left(\sum_{t=1}^{n} x_{p,t}^\star {x_{p,t}^\star}^\top \right)^{-1} \right\| \left\| \sum_{t=1}^{n} x_{p,t}^\star \varepsilon_t^\star \right\|$$

But by lemma 5, the right hand side is $O_p^\star(np^{1/2})O_p^\star(n^{-1})O_p^\star(n^{1/2}p^{1/2})$ a.s., which gives $O_p^\star(n^{1/2}p)$ a.s.. Then, using this result and lemma 4, we have:

$$n^{-1}G_n^\star = \hat{\pi}_n(1)\frac{1}{n}\sum_{t=1}^{n} w_{t-1}^\star \varepsilon_t^\star + o_p^\star(1) \text{ a.s.}$$

We can further say that

$$n^{-2}H_n^\star = \hat{\pi}_n(1)^2\frac{1}{n^2}\sum_{t=1}^{n} {w_{t-1}^\star}^2 + o_p^\star(1) \text{ a.s.}$$

because n^{-2} times the second part of $H_n^\star$ goes to 0 as $n \to \infty$. Therefore, the $\tau_n^\star$ statistic can be seen to be:

$$\tau_n^\star = \frac{\frac{1}{n}\sum_{t=1}^n w_{t-1}^\star \varepsilon_t^\star}{\sigma\left(\frac{1}{n^2}\sum_{t=1}^n {w_{t-1}^\star}^2\right)^{1/2}} + o_p^\star(1) \text{ a.s.}$$

It is then easy to use the results of section 2.2 along with the continuous mapping theorem to deduce that:

$$\frac{1}{n}\sum_{t=1}^n w_{t-1}^\star \varepsilon_t^\star \xrightarrow{d^\star} \int_0^1 W_t dW_t \text{ a.s.}$$

$$\frac{1}{n^2}\sum_{t=1}^n {w_{t-1}^\star}^2 \xrightarrow{d^\star} \int_0^1 W_t^2 dt \text{ a.s.}$$

under assumptions 1 and 4. We can therefore state the following theorem.

Theorem 3. *Under assumptions 1 and 4, we have*

$$\tau_n^\star \xrightarrow{d^\star} \frac{\int_0^1 W_t dW_t}{\left(\int_0^1 W_t^2 dt\right)^{1/2}} \text{ a.s.}$$

which establishes the asymptotic validity of the MASB ADF test.

2.4.2 Consistency of ARMASB ADF tests

It is now very easy to prove the asymptotic validity of ADF tests based on the ARMASB distribution. In order to do this, we make use of the MA(∞) form of the ARMASB DGP (see equation 14 above). In the appendix, it is shown that the results of lemmas A1 to A5 also hold for the ARMASB. Then, we have the following corollaries to lemmas 4 and 5.

Corollary 4. *Under assumptions 1 and 4, we have*

$$\frac{1}{n}\sum_{t=1}^n y_{t-1}^\star \varepsilon_t^\star = \hat\theta_n(1)\frac{1}{n}\sum_{t=1}^n w_{t-1}^\star \varepsilon_t^\star + o_p^\star(1) \text{ a.s.} \tag{27}$$

$$\frac{1}{n^2}\sum_{t=1}^n {y_{t-1}^\star}^2 = \hat\theta_n(1)^2\frac{1}{n^2}\sum_{t=1}^n {w_{t-1}^\star}^2 + o_p^\star(1) \text{ a.s.} \tag{28}$$

Corollary 5. *Under assumptions 1 and 4 we have*

$$\left\|\left(\frac{1}{n}\sum_{t=1}^n x_{p,t}^\star {x_{p,t}^\star}^\top\right)^{-1}\right\| = O_p^\star(1) \text{ a.s.} \tag{29}$$

$$\left\| \sum_{t=1}^{n} x_{p,t}^{\star} y_{t-1}^{\star} \right\| = O_p^{\star}(np^{1/2}) \text{ a.s.} \tag{30}$$

$$\left\| \sum_{t=1}^{n} x_{p,t}^{\star} \varepsilon_t^{\star} \right\| = O_p^{\star}(n^{1/2} p^{1/2}) \text{ a.s.} \tag{31}$$

The proofs may be found in the appendix. Theorem 4 evidently follows.

Theorem 4. *Under assumptions 1 and 4, we have*

$$\tau_n^{\star} \xrightarrow{d^{\star}} \frac{\int_0^1 W_t dW_t}{\left(\int_0^1 W_t^2 dt \right)^{1/2}} \; a.s.$$

This establishes the asymptotic validity of the ARMASB ADF test.

3 Simulations

We now present a set of simulations designed to illustrate the extent to which the proposed MASB and ARMASB schemes improve upon the usual ARSB. Recently, CP 2003 have shown, through Monte Carlo experiments, that the ARSB may allow one to reduce the ADF test's ERP. Such results are generally interpreted as evidence of the presence of asymptotic refinements in the sense of Beran (1987), even though no theoretical proof exists to date. The use of different versions of the block bootstrap has also been suggested, see Paparoditis and Politis (2003), Parker, Paparoditis and Politis (2006) and Swensen (2003). The very complete simulation study of Palm, Smeekes and Urbain (2006) indicates that the ARSB performs slightly better than several versions of the block bootstrap. Therefore, in order to make this section as concise as possible, we only compare the MASB and ARMASB tests to the ARSB and asymptotic ones.

In order to obtain valid test statistics, one must make sure that the parameters used in the construction of the bootstrap DGP are consistent estimates under the null. This is easily achieved in the present case by estimating the appropriate time series model (ie: AR, MA or ARMA) of the first difference of y_t using the maximum likelihood estimator. Bootstrap samples of the first difference process are then built using these parameter estimates and drawings from the residuals. It is also necessary that the bootstrap first difference process be stationary and invertible. We achieve this by constraining our maximum likelihood algorithm to return solutions within the stationarity and invertibility regions. The figures reported below were constructed from 1 000 Monte Carlo samples of 250 realizations from the DGP described below with N(0,1) innovations. The bootstrap tests were conducted using 499 bootstrap samples. The orders (p and q) of all the sieve bootstrap DGPs as well as those of all the ADF regressions were chosen with AIC. We have allowed a maximum order of 8 for the ADF regressions and the AR and MA sieves and of only 2 for the ARMA sieve. Evidently, these choices, which were chiefly made to reduce the computational cost of the experiments, have an impact on the results of the simulations. In particular, allowing for a higher maximum order for the ADF regression and the ARSB probably would have reduced these test's ERP, just like it would have done for the MASB in the simulations presented in figure 2. It is unlikely, however, that it would have affected the good preformances of the ARMASB.

[Figure 1 about here]

Most simulation studies on the overrejection of the ADF test define the first difference process as an MA(1). This is indeed a convenient specification, for it allows one to control the degree of persistence of the series by changing the value of the only MA parameter, θ. It also allows one's analysis to include the near cancellation of the unit root with the MA root which occurs when θ is close to -1. It is in such cases that the ADF test tends to overreject the most. Figure 1 shows the rejection probability (RP), as a function of θ, of the ADF test based on the DF critical values and the ARSB, MASB and ARMASB. The fact that the ARSB has lower ERP than the asymptotic test confirms the findings of CP 2003. However, the MASB and ARMASB seem to be much more robust to the value of the MA parameter. This was, of course, to be expected because the MASB and ARMASB are able to directly estimate the MA part, whereas the ARSB must approximate it with a long autoregression. It is also not surprising that the ARMASB tends to slightly underreject since it is based on an overspecified model. The second DGP we have used is a long MA(33) model with a decaying lag structure function of a parameter θ:

$$\Delta y_t = \varepsilon_t + \theta \left(0.99L + 0.96L^2 + 0.93L^3 + ... + 0.03L^{33}\right) \varepsilon_t, \quad \varepsilon_t \sim N(0,1)$$

where we let θ vary between -0.8 and 0.8. Figure 2 shows the rejection probability of the four tests as a function of θ. The ARMASB test always has lower ERP than the other three. On the other hand, the ARSB is now working better than the MASB. This is not surprising since the DGP looks somewhat like an inverted low order AR process. Similar results, with the MASB outperforming the ARSB, were obtained with an AR(33) DGP. It is worth noting that the average order chosen by AIC, through all values of θ considered, is between 2.6 and 5.5 for the AR sieve and between 4.6 and 7.9 for the MA sieve, while it is always around 1.2 for both parts of the ARMA sieve. The ARMASB is thus able to provide greater accuracy then the other two with a more parsimonious bootstrap DGP.

[Figure 2 about here]

One might wonder whether the accuracy gains of the ARMASB under the null come at any cost in terms of power. As an illustration, we have looked at the power function of the four tests in the MA(1) model when $\theta = -0.4$. It can be seen from figure 3 that it is indeed the case. However, the difference between the RP of the standard ADF and that of the MASB and ARMASB is roughly constant around 0.05 for the set of alternative hypothesis considered. It therefore appears likely that, where we to use size-corrected critical values, the power loss would vanish.

[Figure 3 about here]

Finally, it should be noted that unreported simulations with smaller sample sizes have provided similar results, thus indicating that our results are robust in that respect as well.

4 Conclusion

Since its introduction, the sieve bootstrap has only been used with AR approximations. We propose to use MA and ARMA approximations. In the unit root case, we provide invariance principles that may be used to derive asymptotic properties for the tests performed using these sieve bootstraps. We then show that ADF unit root tests based on them are asymptotically valid. Similar results were provided for the usual ARSB by Park (2002) and Chang and Park (2003).

Our simulations confirm existing results in the literature that using a sieve bootstrap to carry out ADF tests may lead to more accurate testing than using the DF distribution's critical values. They also indicate that using MA or ARMA sieves instead of AR ones may allow for a further accuracy gain. The power of these two new test procedures was found to be quite acceptable. Overall, the ARMASB test appears to be the best choice because of its robustness to the underlying DGP and its parsimony.

A Appendix: Mathematical Proofs

Proof of Lemma 1.

Consider the following expression for the AR sieve bootstrap model taken from Park (2002, equation 19):

$$u_t = \alpha_{f,1} u_{t-1} + \alpha_{f,2} u_{t-2} + ... + \alpha_{f,f} u_{t-f} + e_{f,t} \tag{32}$$

where the coefficients $\alpha_{f,1}$ are pseudo-true values defined so that the equality holds and the $e_{f,t}$ are uncorrelated with the u_{t-k}, $k = 0, 1, 2, ...r$. Using the binding functions of GZW (1994), we define

$$\pi_{q,1} = \alpha_{f,1}$$
$$\pi_{q,2} = \alpha_{f,2} + \alpha_{f,1}\pi_{q,1}$$
$$\vdots$$
$$\pi_{q,q} = \alpha_{f,q} + \alpha_{f,q-1}\pi_{q,1} + ... + \alpha_{f,1}\pi_{q,q-1}$$

to be the moving average parameters deduced from the pseudo-true parameters of the AR model (32). Let $\hat{\alpha}_{f,k}$ be the OLS estimator of these parameters. Also, let $\hat{\pi}_{q,k}$ be the GZW 1994 estimator of the $\pi_{q,k}$. We can use this estimator without any loss of generality because it is consistent. Thus,

$$|\hat{\pi}_{q,k} - \pi_k| = \left| \sum_{j=0}^{k-1} (\hat{\alpha}_{f,k-j}\hat{\pi}_{q,j} - \alpha_{k-j}\pi_j) \right| \tag{33}$$

The result then follows from Park (2002, lemma 3.1) who, building on results of Baxter (1962) and Hannan and Kavalieris (1986), shows that:

$$\max_{1 \le k \le f} |\hat{\alpha}_{f,k} - \alpha_k| = O\left((\log n/n)^{1/2} \right) + o(f^{-s}) \text{ a.s.}$$

Indeed, consider, for example, $k = 3$. Then,

$$|\hat{\pi}_{q,3} - \pi_3| = \left| \hat{\alpha}_{f,1}^3 + 2\hat{\alpha}_{f,1}\hat{\alpha}_{f,2} + \hat{\alpha}_{f,3} - \alpha_1^3 - 2\alpha_1\alpha_2 - \alpha_3 \right|$$

using the triangle inequality:

$$|\hat{\pi}_{q,3} - \pi_3| \le |\hat{\alpha}_{f,1}^3 - \alpha_1^3| + |2(\hat{\alpha}_{f,1}\hat{\alpha}_{f,2} - \alpha_1\alpha_2)| + |\hat{\alpha}_{f,3} - \alpha_3|$$

By lemma 3.1 of Park (2002), every element of the right hand side is $o(1)$ a.s.. This is obviously true for all k. The other parts of the lemma follow immediately. ■

Proof of Lemma 2.

First, note that $n^{1-r/2}\mathrm{E}^\star |\varepsilon_t^\star|^r = n^{1-r/2} \left(\frac{1}{n} \sum_{t=1}^n \left| \hat{\varepsilon}_{q,t} - \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_{q,t} \right|^r \right)$ because the bootstrap errors are drawn from the series of recentered residuals from the MA(q) model. Therefore, what must be shown is that

$$n^{1-r/2} \left(\frac{1}{n} \sum_{t=1}^n \left| \hat{\varepsilon}_{q,t} - \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_{q,t} \right|^r \right) \xrightarrow{a.s.} 0$$

as $n \to \infty$. If we add and subtract ε_t and $\varepsilon_{q,t}$ (which was defined in equation 6) inside the absolute value operator, we obtain:

$$n^{1-r/2} \left(\frac{1}{n} \sum_{t=1}^n \left| \hat{\varepsilon}_{q,t} - \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_{q,t} \right|^r \right) \le c\,(A_n + B_n + C_n + D_n)$$

where c is a constant and

$A_n = \frac{1}{n}\sum_{t=1}^{n} |\varepsilon_t|^r$

$B_n = \frac{1}{n}\sum_{t=1}^{n} |\varepsilon_{q,t} - \varepsilon_t|^r$

$C_n = \frac{1}{n}\sum_{t=1}^{n} |\hat{\varepsilon}_{q,t} - \varepsilon_{q,t}|^r$

$D_n = \left|\frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}_{q,t}\right|^r$

To get the desired result, one must show that $n^{1-r/2}$ times A_n, B_n, C_n and D_n each go to 0 almost surely.

1. $n^{1-r/2}A_n \overset{a.s.}{\to} 0$**.**

This holds by the strong law of large numbers which states that $A_n \overset{a.s.}{\to} E\left(|\varepsilon_t|^r\right)$ which has been assumed to be finite. Since $r > 4$, $1 - r/2 < -1$ from which the result follows.

2. $n^{1-r/2}B_n \overset{a.s.}{\to} 0$**.**

This is proved by showing that

$$E\left(|\varepsilon_{q,t} - \varepsilon_t|^r\right) = o(q^{-rs}) \tag{34}$$

holds uniformly in t and where s is as specified in assumption 1 part b. From equation (6) we have

$$\varepsilon_{q,t} = u_t - \sum_{k=1}^{q} \pi_k \varepsilon_{q,t-k}.$$

Writing this using an infinite AR form:

$$\varepsilon_{q,t} = u_t - \sum_{k=1}^{\infty} \ddot{\alpha}_k u_{t-k}$$

where the parameters $\ddot{\alpha}_k$ are functions of the first q true parameters π_k in the usual manner (see proof of lemma 1). We also have:

$$\varepsilon_t = u_t - \sum_{k=1}^{\infty} \pi_k \varepsilon_{t-k}.$$

which we also write in AR(∞) form:

$$\varepsilon_t = u_t - \sum_{k=1}^{\infty} \alpha_k u_{t-k}.$$

Evidently, $\alpha_k = \ddot{\alpha}_k$ for all $k = 1, ..., q$. Subtracting the second of these two expressions to the first we obtain:

$$\varepsilon_{q,t} - \varepsilon_t = \sum_{k=q+1}^{\infty} (\alpha_k - \ddot{\alpha}_k)\, u_{t-k}. \tag{35}$$

Using Minkowski's inequality, the triangle inequality and the stationarity of u_t,

$$E\left(|\varepsilon_{q,t} - \varepsilon_t|^r\right) \leq E\left(|u_t|^r\right) \left(\sum_{k=q+1}^{\infty} |(\alpha_k - \ddot{\alpha}_k)|\right)^r. \tag{36}$$

The second element of the right hand side can be rewritten as

$$\left(\sum_{k=q+1}^{\infty} \left|\sum_{\ell=1}^{q} \pi_\ell \alpha_{k-\ell} - \sum_{\ell=1}^{k} \pi_\ell \alpha_{k-\ell} - \pi_k\right|\right)^r = \left(\sum_{k=q+1}^{\infty} \left|-\sum_{\ell=1}^{k-q} \pi_\ell \alpha_{k-\ell} - \pi_k\right|\right)^r.$$

Equation (34) therefore follows from assumptions 1 b and the rate given to q in assumption 2.

3. $n^{1-r/2} C_n \overset{a.s.}{\to} 0$

We start from the AR(∞) expression for the residuals to be resampled:

$$\hat{\varepsilon}_{q,t} = u_t - \sum_{k=1}^{\infty} \hat{\alpha}_{q,k} u_{t-k} \tag{37}$$

where $\hat{\alpha}_{q,k}$ denotes the parameters corresponding to the estimated MA(q) parameters $\hat{\pi}_{q,k}$ through the analytical binding functions. Then, adding and subtracting $\sum_{k=1}^{\infty} \alpha_{q,k} u_{t-k}$, where the parameters $\alpha_{q,k}$ are function of the $\pi_{q,k}$ as defined in the proof of lemma 1, and using once more the AR(∞) form of (6), equation (37) becomes:

$$\hat{\varepsilon}_{q,t} = \varepsilon_{q,t} - \sum_{k=1}^{\infty} (\hat{\alpha}_{q,k} - \alpha_{q,k}) u_{t-k} - \sum_{k=1}^{\infty} (\alpha_{q,k} - \ddot{\alpha}_k) u_{t-k}. \tag{38}$$

It then follows that

$$|\hat{\varepsilon}_{q,t} - \varepsilon_{q,t}|^r \leq c \left(\left|\sum_{k=1}^{\infty} (\hat{\alpha}_{q,k} - \alpha_{q,k}) u_{t-k}\right|^r + \left|\sum_{k=1}^{\infty} (\alpha_{q,k} - \ddot{\alpha}_k) u_{t-k}\right|^r \right)$$

for $c = 2^{r-1}$. Let us define

$$C_{1n} = \frac{1}{n} \sum_{t=1}^{n} \left|\sum_{k=1}^{\infty} (\hat{\alpha}_{q,k} - \alpha_{q,k}) u_{t-k}\right|^r$$

$$C_{2n} = \frac{1}{n} \sum_{t=1}^{n} \left|\sum_{k=1}^{\infty} (\alpha_{q,k} - \ddot{\alpha}_k) u_{t-k}\right|^r.$$

Then showing that $n^{1-r/2}C_{1n} \overset{a.s.}{\to} 0$ and $n^{1-r/2}C_{2n} \overset{a.s.}{\to} 0$ will give us our result. First, let us note that C_{1n} is majorized by:

$$\left(\max_{1\leq k\leq\infty} |\hat{\alpha}_{q,k} - \alpha_{q,k}|^r\right) \frac{1}{n}\sum_{t=1}^{n}\sum_{k=1}^{\infty} |u_{t-k}|^r . \tag{39}$$

By our lemma 1 and equation (20) of Park (2002), we have

$$\max_{1\leq k\leq\infty} |\hat{\alpha}_{q,k} - \alpha_{q,k}| = O\left((\log n/n)^{1/2}\right) \text{ a.s.}$$

Thus, the first part of (39) goes to 0. Also, the second part is bounded away from infinity by a law of large numbers and equation (25) in Park (2002). This proves the first result. Applying Minkowski's inequality to the absolute value part of C_{2n},

$$E\left(\left|\sum_{k=1}^{\infty}\left(\alpha_{q,k} - \ddot{\alpha}_k\right)u_{t-k}\right|^r\right) \leq E\left(|u_t|^r\right)\left(\sum_{k=1}^{\infty}|\alpha_{q,k} - \ddot{\alpha}_k|\right)^r \tag{40}$$

the right hand side of which goes to 0 by the boundedness of $E\left(|u_t|^r\right)$, the definition of the $\ddot{\alpha}$, lemma 1 and equation (21) of Park (2002), where it is shown that $\sum_{k=1}^{p}|\alpha_{p,k} - \alpha_k| = o(p^{-s})$ for some $p \to \infty$, which implies a similar result between the $\pi_{q,k}$ and π_k, which in turn implies a similar result between the $\alpha_{q,k}$ and $\ddot{\alpha}_k$.

4. $n^{1-r/2}D_n \overset{a.s.}{\to} 0$

In order to prove this, we show that

$$\frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_{q,t} = \frac{1}{n}\sum_{t=1}^{n}\varepsilon_{q,t} + o(1) \text{ a.s.} = \frac{1}{n}\sum_{t=1}^{n}\varepsilon_t + o(1) \text{ a.s.}$$

Recalling equations (35) and (38), this is true if

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{k=q+1}^{\infty}\left(\alpha_k - \ddot{\alpha}_k\right)u_{t-k} \overset{a.s.}{\to} 0 \tag{41}$$

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{k=1}^{\infty}(\alpha_{q,k} - \ddot{\alpha}_k)u_{t-k} \overset{a.s.}{\to} 0 \tag{42}$$

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{k=1}^{\infty}(\hat{\alpha}_{q,k} - \alpha_{q,k})u_{t-k} \overset{a.s.}{\to} 0 \tag{43}$$

where the first expression serves to prove the second asymptotic equality and the other two serve to prove the first asymptotic equality. Proving those 3 results requires some work. Just like Park (2002), p.485, let us define

$$S_n(i,j) = \sum_{t=1}^{n}\varepsilon_{t-i-j}$$

and

$$T_n(i) = \sum_{t=1}^{n} u_{t-i}$$

so that

$$T_n(i) = \sum_{j=0}^{\infty} \pi_j S_n(i,j)$$

and remark that by Doob's inequality,

$$\left| \max_{1 \leq m \leq n} S_m(i,j) \right|^r \leq z \left| S_n(i,j) \right|^r$$

where $z = 1/(1-\frac{1}{r})$. Taking expectations and applying Burkholder's inequality,

$$E\left(\max_{1 \leq m \leq n} \left| S_m(i,j) \right|^r \right) \leq c_1 z E\left(\left| \sum_{t=1}^{n} \varepsilon_t^2 \right|^{r/2} \right)$$

where c_1 is a constant depending only on r. By the law of large numbers, the right hand side is equal to $c_1 z (n\sigma^2)^{r/2}$=$c_1 z (n^{r/2}\sigma^r)$. Thus, we have

$$E\left(\max_{1 \leq m \leq n} \left| S_m(i,j) \right|^r \right) \leq c n^{r/2}$$

uniformly over i and j, where $c = c_1 z \sigma^r$. Define

$$L_n = \sum_{k=q+1}^{\infty} (\alpha_k - \ddot{\alpha}_k)\, T_n(k) = \sum_{k=q+1}^{\infty} (\alpha_k - \ddot{\alpha}_k) \sum_{t=1}^{n} u_{t-k}.$$

It must therefore follow that

$$\left[E\left(\max_{1 \leq m \leq n} |L_m|^r \right) \right]^{1/r} \leq \sum_{k=q+1}^{\infty} |(\alpha_k - \ddot{\alpha}_k)| \left[E\left(\max_{1 \leq m \leq n} |T_m(k)|^r \right) \right]^{1/r}$$

$$\leq \sum_{k=q+1}^{\infty} |(\alpha_k - \ddot{\alpha}_k)|\, c n^{1/2}$$

where the constant c is redefined accordingly. But

$$\sum_{k=q+1}^{\infty} |(\alpha_k - \ddot{\alpha}_k)| = o(q^{-s})$$

by assumption 1 b and the construction of the $\ddot{\alpha}_k$. Thus,

$$E\left(\max_{1 \leq m \leq n} |L_m|^r \right) \leq c q^{-rs} n^{r/2}.$$

Then it follows from the result in Moricz (1976, theorem 6), that for any $\delta > 0$,

$$L_n = o\left(q^{-s} n^{1/2} (\log n)^{1/r} (\log \log n)^{(1+\delta)/r}\right) = o(n) \text{ a.s..}$$

This last equation proves (41). Now, if we let

$$M_n = \sum_{k=1}^{\infty} (\alpha_{q,k} - \ddot{\alpha}_k) T_n(k) = \sum_{k=1}^{\infty} (\alpha_{q,k} - \ddot{\alpha}_k) \sum_{t=1}^{n} u_{t-k},$$

then, we find by the same devise that

$$\left[E\left(\max_{1 \le m \le n} |M_m|^r\right)\right]^{1/r} \le \sum_{k=1}^{\infty} |(\alpha_{q,k} - \ddot{\alpha}_k)| \left[E\left(\max_{1 \le m \le n} |T_m(k)|^r\right)\right]^{1/r}$$

the right hand side of which is smaller than or equal to $cq^{-s}n^{1/2}$, see the discussion under equation (40). Consequently, using Moricz's result once again, we have $M_n = o(n)$ a.s. and (42) is proved. Finally, define

$$N_n = \sum_{k=1}^{\infty} (\hat{\alpha}_{q,k} - \alpha_{q,k}) T_n(k) = \sum_{k=1}^{\infty} (\hat{\alpha}_{q,k} - \alpha_{q,k}) \sum_{t=1}^{n} u_{t-k}$$

further, let

$$Q_n = \sum_{k=1}^{\infty} \left| \sum_{t=1}^{n} u_{t-k} \right|.$$

Then, the larger element of N_n is $Q_n \max_{1 \le k \le \infty} |\hat{\alpha}_{q,k} - \alpha_{q,k}|$. By assumption 1 b and the result under (39), we know that the second part goes to 0. Then, using again Doob's and Burkholder's inequalities,

$$E\left(\max_{1 \le m \le n} |Q_m|^r\right) \le cq^r n^{r/2}.$$

Therefore, for any $\delta > 0$,

$$Q_n = o\left(q n^{1/2} (\log n)^{1/r} (\log \log n)^{(1+\delta)/r}\right) \text{ a.s.}$$

and

$$N_n = O\left((\log n/n)^{1/2}\right) Q_n = o\left(q (\log n)^{(r+2)/2r} (\log \log n)^{(1+\delta)/r}\right) = o(n).$$

Hence, equation (43) is proved. The proof of the lemma is complete. ■

Proof of Lemma 3.

We begin by noting that

$$P^\star\left(\max_{1 \le t \le n} \left| n^{-1/2} \tilde{u}_t^\star \right| > \delta\right) \le \sum_{t=1}^{n} P^\star\left(\left| n^{-1/2} \tilde{u}_t^\star \right| > \delta\right)$$

$$= nP^\star\left(\left|n^{-1/2}\tilde{u}_t^\star\right| > \delta\right)$$
$$\leq (1/\delta^r)n^{1-r/2}E^\star\left(\left|\tilde{u}_t^\star\right|^r\right)$$

where the first inequality is trivial, the equality follows from the stationarity of $\tilde{u}_t^\star$ conditional on the realization of $\{\hat{\varepsilon}_{q,t}\}$ and the last inequality is an application of the Tchebyshev inequality. Recall that

$$\tilde{u}_t^\star = \sum_{k=1}^{q}\left(\sum_{i=k}^{q}\hat{\pi}_{q,i}\right)\varepsilon_{t-k+1}^\star.$$

Then, by Minkowski's inequality and assumption 1, we have:

$$E^\star\left(\left|\tilde{u}_t^\star\right|^r\right) \leq \left(\sum_{k=1}^{q}k\left|\hat{\pi}_{q,k}\right|\right)^r E^\star\left(\left|\varepsilon_t^\star\right|^r\right)$$

But by lemma 1, the estimates $\hat{\pi}_{q,k}$ are consistent for π_k. Hence, by assumption 1, the first part must be bounded as $n \to \infty$. Also, we have shown in lemma 2 that $n^{1-r/2}E^\star\left(\left|\varepsilon_t^\star\right|^r\right) \overset{a.s.}{\to} 0$. The result thus follows. ∎

Proof of Theorem 1.

The result follows directly from combining lemmas 1, 2 and 3. ∎

Proof of Corollary 1

The proof of this corollary is a direct extension of the proof of lemma 1. It suffices to point out the fact that the $\hat{\theta}_{\ell,k}$ can be written as functions of the $\hat{\pi}_{\ell,j}$ and $\hat{\alpha}_{\ell,j}$, which can themselves be expressed as functions of the parameters of a long autoregression using the binding functions of GZW (1997). We know, by lemma 3.1 of Park (2002), that the estimated parameters of this long autoregression are consistent estimators of the parameters of the AR(∞) form of u_t. The corollary follows by arguments similar to those used in the proof of lemma 1. ∎

Proof of corollary 2.

Recall that $\ell = p + q$ and assume that $\ell \to \infty$ at the rate specified in assumption 3. Then, for any value of $p \in [0, \infty)$ and $q \to \infty$, the stated result follows from lemma 2 and corollary 1. To see this, define ν_t as being the original process u_t with an AR(p) part filtered out using the p consistent parameter estimates. Then, ν_t is an invertible general linear process to which we can apply lemma 2.

By the same logic, letting $q \in [0, \infty)$ and $p \to \infty$, and defining ν_t as being the original process with the MA(q) part filtered out, it is easily seen that the result of lemma 3.2 in Park (2002) can be applied. Since these two situations allow us to handle every possible cases, the result is shown. ∎

Proof of corollary 3

Identical to corollary 2 except that we use lemma 3 for the case where $q \to \infty$ and the proof of theorem 3.3 in Park (2002) for the case where $p \to \infty$. ∎

Proof of Theorem 2.

The result follows directly from combining corollaries 1 to 3. ∎

Lemma A 1. *Under assumptions 1 and 2, $e_t = \varepsilon_t^\star + o(1)$ a.s.*

Proof of Lemma A1.

Let us first rewrite the ADF regression (21) under the null as follows:

$$\Delta y_t^\star = \sum_{k=1}^{p} \alpha_{p,k} \left(\sum_{j=k+1}^{k+q} \hat{\pi}_{q,j-k} \varepsilon_{t-j}^\star + \varepsilon_{t-k}^\star \right) + e_t$$

where we have substituted the MASB DGP for $\Delta y_{t-k}^\star$. This can be rewritten as

$$\Delta y_t^\star = \sum_{i=1}^{p} \sum_{j=0}^{q} \alpha_{p,i} \hat{\pi}_{q,j} \varepsilon_{t-i-j}^\star + e_t. \tag{44}$$

where $\hat{\pi}_{0,j}$ is constrained equal 1. Then, from equations (44) and (7),

$$e_t = \sum_{j=1}^{q} \hat{\pi}_{q,j} \varepsilon_{t-j}^\star + \varepsilon_t^\star - \sum_{i=1}^{p} \sum_{j=0}^{q} \alpha_{p,i} \hat{\pi}_{q,j} \varepsilon_{t-i-j}^\star.$$

We can write $e_t = E_t + F_t + \varepsilon_t^\star$, where

$E_t = (\hat{\pi}_{q,1} - \alpha_1)\varepsilon_{t-1}^\star + (\hat{\pi}_{q,2} - \alpha_{p,1}\hat{\pi}_{q,1} - \alpha_{p,2})\varepsilon_{t-2}^\star + \ldots + (\hat{\pi}_{q,q} - \alpha_{p,1}\hat{\pi}_{q,q-1} - \ldots - \alpha_{p,q})\varepsilon_{t-q}^\star$

and

$$F_t = \sum_{j=q+1}^{p+q} \varepsilon_{t-j}^\star \left(\sum_{i=0}^{q} \hat{\pi}_{q,i} \alpha_{p,j-i} \right)$$

where $\hat{\pi}_0 = 1$ and $\alpha_{p,i} = 0$ whenever $i > p$. First, we note that the coefficients appearing in E_t are the binding functions of GZW 1994. Thus, by lemma 1 of the present paper and lemma 3.1 of Park (2002), E_t goes to 0 almost surely. On the other hand, taking F_t and applying Minkowski's inequality to it gives:

$$E^\star\left(|F_t|^r\right) \leq E^\star\left(|\varepsilon_t^\star|^r\right) \left(\sum_{j=q+1}^{p+q} \sum_{i=0}^{q} |\hat{\pi}_{q,i} \alpha_{p,j-i}| \right)^r.$$

But for each j, $\sum_{i=0}^{q} \pi_{q,i} \alpha_{p,j-i} = -\alpha_{p,j}$. Hence, by lemma 1.

$$E^\star\left(|F_t|^r\right) \leq E^\star\left(|\varepsilon_t^\star|^r\right) \left(\sum_{j=q+1}^{p+q} |\alpha_{p,j}| \right)^r \text{ a.s.}$$

But, as p and q go to infinity with $p > q$, $\sum_{j=q+1}^{p+q}(|\alpha_j|)^r = o(q^{-rs})$ a.s. under assumption 1. The result therefore follows since $E^\star\left(|\varepsilon_t^\star|^r\right)$ is finite. The proof for the ARMASB follows along the same line and we therefore omit it. ■

Lemma A2 (CP lemma A1). *Under assumptions 1 and 4, we have* $\sigma_\star^2 \overset{a.s.}{\to} \sigma^2$ *and* $\Gamma_0^\star \overset{a.s.}{\to} \Gamma_0$ *as* $n \to \infty$ *where* $E^\star\left(|\varepsilon_t^\star|^2\right) = \sigma_\star^2$ *and* $E^\star\left(|u_t^\star|^2\right) = \Gamma_0^\star$

Proof of Lemma A2.

Consider the MASB DGP (7) once more:

$$u_t^\star = \sum_{k=1}^{q} \hat{\pi}_{q,k}\varepsilon_{t-k}^\star + \varepsilon_t^\star \tag{45}$$

Under assumption 1 and given lemma 1, this process admits an AR(∞) representation:

$$u_t^\star + \sum_{k=1}^{\infty} \hat{\psi}_{q,k}u_{t-k}^\star = \varepsilon_t^\star \tag{46}$$

where we write the $\hat{\psi}_{q,k}$ parameters with a hat and a subscript q to emphasize that they come from the estimation of a finite order MA(q) model. We can rewrite equation (46) as follows:

$$u_t^\star = -\sum_{k=1}^{\infty} \hat{\psi}_{q,k}u_{t-k}^\star + \varepsilon_t^\star. \tag{47}$$

Multiplying by $u_t^\star$ and taking expectations under the bootstrap DGP, we obtain

$$\Gamma_0^\star = -\sum_{k=1}^{\infty} \hat{\psi}_{q,k}\Gamma_k^\star + \sigma_\star^2.$$

dividing both sides by $\Gamma_0^\star$ and rearranging,

$$\Gamma_0^\star = \frac{\sigma_\star^2}{1 + \sum_{k=1}^{\infty} \hat{\psi}_{q,k}\rho_k^\star} \tag{48}$$

where $\rho_k^\star$ are the autocorrelations of the bootstrap process. Note that these are functions of the parameters $\hat{\psi}_{q,k}$ and that it can easily be shown that they satisfy the homogeneous system of linear differential equations described as:

$$\rho_h^\star + \sum_{k=1}^{\infty} \hat{\psi}_{q,k}\rho_{h-k}^\star = 0$$

for all $h > 0$.On the other hand, we have:

$$u_t = \sum_{k=1}^{q} \hat{\pi}_{q,k}\hat{\varepsilon}_{q,t-k} + \hat{\varepsilon}_{q,t} \tag{49}$$

which has an AR(∞) representation:

$$u_t = -\sum_{k=1}^{\infty} \hat{\psi}_{q,k} u_{t-k} + \hat{\varepsilon}_{q,t}.$$

where the $\hat{\psi}_{q,k}$ are exactly the same as in equation (47). This yields

$$\Gamma_{0,n} = \frac{\hat{\sigma}_n^2}{1 + \sum_{k=1}^{\infty} \hat{\psi}_{q,k} \rho_{k,n}} \tag{50}$$

where $\Gamma_{0,n}$, is the sample autocovariance of u_t when we have n observations, $\hat{\sigma}_n^2 = (1/n) \sum_{i=1}^{n} \hat{\varepsilon}_t^2$ and $\rho_{k,n}$ is the k^{th} autocorrelation of u_t. Since the autocorrelation parameters are the same in equations (48) and (50), we can write:

$$\Gamma_0^\star = (\sigma_\star^2 / \hat{\sigma}_n^2) \Gamma_{0,n}.$$

The strong law of large numbers implies that $\Gamma_{0n} \stackrel{a.s.}{\to} \Gamma_0$. Therefore, we only need to show that $\sigma_\star^2 / \hat{\sigma}_n^2 \stackrel{a.s.}{\to} 1$. By lemma 1, $\hat{\sigma}_n^2 \stackrel{a.s.}{\to} \sigma^2$. Also, since the $\varepsilon_t^\star$ are drawn from the EDF of $(\hat{\varepsilon}_{q,t} - (1/n) \sum_{t=1}^{n} \hat{\varepsilon}_{q,t})$,

$$\sigma_\star^2 = \frac{1}{n} \sum_{t=1}^{n} \left(\hat{\varepsilon}_{q,t} - \frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}_{q,t} \right)^2.$$

The independence of the $\varepsilon_t^\star$ implies that:

$$\sigma_\star^2 = \hat{\sigma}_n^2 + \left(\frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}_{q,t} \right)^2. \tag{51}$$

But we have shown in lemma 2 that $((1/n) \sum_{t=1}^{n} \hat{\varepsilon}_{q,t})^2 = o(1)$ a.s. (to see this, take the result for $n^{1-r/2} D_n$ with $r = 4$). Thus, we have $\sigma_\star^2 \stackrel{a.s.}{\to} \hat{\sigma}_n^2$, and hence, $\sigma_\star^2 / \hat{\sigma}_n^2 \stackrel{a.s.}{\to} 1$. It therefore follows that $\Gamma_0^\star \stackrel{a.s.}{\to} \Gamma_0$. On the other hand, $\sigma_\star^2 \stackrel{a.s.}{\to} \hat{\sigma}_n^2$ implies $\sigma_\star^2 \stackrel{a.s.}{\to} \sigma^2$.

To extend this result the ARMASB case, it suffices to replace equations (45) and (49) by their ARMA counterparts and to use the corollaries to lemmas 1 and 2. ∎

Lemma A3 (CP lemma A2, Berk, theorem 1, p. 493). *Let f and $f^\star$ be the spectral densities of u_t and $u_t^\star$ respectively. Then, under assumptions 1 and 4,*

$$\sup_{\lambda} |f^\star(\lambda) - f(\lambda)| = o(1) \text{ a.s.}$$

Also, letting Γ_k and $\Gamma_k^\star$ be the autocovariance functions of u_t and $u_t^\star$ respectively, we have

$$\sum_{k=-\infty}^{\infty} \Gamma_k^\star = \sum_{k=-\infty}^{\infty} \Gamma_k + o(1) \text{ a.s.}$$

Proof of Lemma A3.

We first derive the result for the MASB. The spectral density of the bootstrap is

$$f^{\star}(\lambda) = \frac{\sigma_{\star}^2}{2\pi}\left|1 + \sum_{k=1}^{q} \hat{\pi}_{q,k} e^{ik\lambda}\right|^2 .$$

Further, let us define

$$f(\hat{\lambda}) = \frac{\hat{\sigma}_n^2}{2\pi}\left|1 + \sum_{k=1}^{q} \hat{\pi}_{q,k} e^{ik\lambda}\right|^2$$

and recall that

$$\sigma_{\star}^2 = \frac{1}{n}\sum_{t=1}^{n}\left[\hat{\varepsilon}_{q,t} - \frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_{q,t}\right]^2 .$$

From lemma 2 (proof of the 4^{th} part) and lemma A2 (equation 51), we have

$$\sigma_{\star}^2 = \frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_{q,t}^2 + o_p(1).$$

Thus,

$$f^{\star}(\lambda) = f(\hat{\lambda}) + o_p(1).$$

Therefore, the desired result follows if we show that

$$\sup_{\lambda}\left|f(\hat{\lambda}) - f(\lambda)\right| = o(1) \text{ a.s.}$$

Let $f_n(\lambda)$ be the spectral density function evaluated at the pseudo-true parameters:

$$f_n(\lambda) = \frac{\sigma_n^2}{2\pi}\left|1 + \sum_{k=1}^{q} \pi_{q,k} e^{ik\lambda}\right|^2$$

where σ_n^2 is the minimum value of

$$\int_{-\pi}^{\pi} f(\lambda)\left|1 + \sum_{k=1}^{q} \pi_{q,k} e^{ik\lambda}\right|^{-2} d\lambda$$

and $\sigma_n^2 \to \sigma^2$ as shown in Baxter (1962). Obviously,

$$\sup_{\lambda}\left|f(\hat{\lambda}) - f_n(\lambda)\right| = o(1) \text{ a.s.}$$

by lemma 1 and equation (20) of Park (2002). Also,

$$\sup_{\lambda}\left|f_n(\lambda) - f(\lambda)\right| = o(1) \text{ a.s.}$$

where

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left|1 + \sum_{k=1}^{q} \pi_k e^{ik\lambda}\right|^2.$$

by the same argument we used at the end of part 3 of the proof of lemma 2. The first part of the present lemma therefore follows. If we consider that

$$\sum_{-\infty}^{\infty} \Gamma_k = 2\pi f(0) \quad \text{and} \quad \sum_{-\infty}^{\infty} \Gamma_k^\star = 2\pi f^\star(0)$$

the second part follows directly.

For the ARMASB, the spectral density is

$$f^\star(\lambda) = \frac{\sigma_\star^2}{2\pi}\left|1 + \sum_{k=1}^{\infty} \hat{\theta}_{\ell,k} e^{ik\lambda}\right|^2.$$

where $\hat{\theta}_{\ell,k}$ is the ℓ^{th} parameter of the MA(∞) form of the ARMASB DGP. Since lemmas 1, 2 and A2 also hold for the ARMASB, the proof follows the same line. Specifically, it is easy to see that the result $n^{1-r/2}D_n \overset{a.s.}{\to} 0$ in part four of lemma 2, which is of crucial importance, holds if we replace the parameters of the AR(∞) form of the MASB by those corresponding to the AR(∞) form of the ARMASB. ∎

Lemma A4 (CP lemma A3). *Under assumptions 1 and 4, we have*

$$E^\star\left(|\varepsilon_t^\star|^4\right) = O(1) \text{ a.s.}$$

Proof of Lemma A4.

From the proof of lemma 2, we have $E^\star\left(|\varepsilon_t^\star|^4\right) \leq c\,(A_n + B_n + C_n + D_n)$ where c is a constant. The relevant results are:

1. $A_n = O(1)$ a.s.
2. $E(B_n) = o(q^{-rs})$ (equation 34)
3. $C_n \leq 2^{r-1}(C_{1n} + C_{2n})$
   where $C_{1n} = o(1)$ a.s. (equation above 40)
   and $E(C_{2n}) = o(q^{-rs})$ (equation 40)
4. $D_n = o(1)$ a.s.

Under assumption 4, we have that $B_n = o(1)$ a.s. and $C_{2n} = o(1)$ a.s. because $o(q^{-rs}) = o((cn^k)^{-rs}) = o(n^{-krs}) = o(n^{-1-\delta})$ for $\delta > 0$. ∎

Lemma A5 (CP lemma A4, Berk proof of lemma 3). *Define*

$$M_n^\star(i,j) = E^\star \left(\sum_{t=1}^{n} \left(u_{t-i}^\star u_{t-j}^\star \right) - \Gamma_{i-j}^\star \right)^2 .$$

Then, under assumptions 1 and 4, we have $M_n^\star(i,j) = O(n)$ *a.s.*

Proof of Lemma A5.

For general linear models, Hannan (1960, p. 39) and Berk (1974, p. 491) have shown that

$$M_n^\star(i,j) \leq n \left[2 \sum_{k=-\infty}^{\infty} \Gamma_k^\star + |K_4^\star| \left(\sum_{k=0}^{\infty} \hat{\pi}_{q,k}^2 \right)^2 \right]$$

for all i and j and where $K_4^\star$ is the fourth cumulant of $\varepsilon_t^\star$. Since our MASB and ARMASB are general linear models, this result applies here. But $K_4^\star$ can be written as a polynomial of degree 4 in the first 4 moments of $\varepsilon_t^\star$. Therefore, $|K_4^\star|$ must be $O(1)$ a.s. by lemma A4. The result now follows from lemma A3 and the fact that $\sum_{k=-\infty}^{\infty} \Gamma_k = O(n)$. ∎

Before going on, it is appropriate to note that the proofs of lemmas 4 and 5 are almost identical to the proofs of lemma 3.2 and 3.3 of CP (2003). We present them here for the sake of completeness. For ease of notation and to be consistent with the preceeding proofs, we define $\Delta y_t \equiv u_t$ and $\Delta y_t^\star \equiv u_t^\star$.

Proof of Lemma 4.

First, we prove equation (22). Using the Beveridge-Nelson decomposition of $u_t^\star$ and the fact that $y_t^\star = \sum_{k=1}^{t} u_k^\star$, we can write:

$$\frac{1}{n} \sum_{t=1}^{n} y_t^\star \varepsilon_t^\star = \hat{\pi}_n(1) \frac{1}{n} \sum_{t=1}^{n} w_{t-1}^\star \varepsilon_t^\star + \tilde{u}_0^\star \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t^\star - \frac{1}{n} \sum_{t=1}^{n} \tilde{u}_{t-1}^\star \varepsilon_t^\star .$$

Therefore, to prove the first result, it suffices to show that

$$E^\star \left(\tilde{u}_0^\star \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t^\star - \frac{1}{n} \sum_{t=1}^{n} \tilde{u}_{t-1}^\star \varepsilon_t^\star \right) = o(1) \ a.s. \tag{52}$$

Since the $\varepsilon_t^\star$ are *iid* by construction, we have:

$$E^\star \left(\sum_{t=1}^{n} \varepsilon_t^\star \right)^2 = n\sigma_\star^2 = O(n) \ a.s \tag{53}$$

and

$$E^\star\left(\sum_{t=1}^{n}\tilde{u}^\star_{t-1}\varepsilon^\star_t\right)^2 = n\sigma^2_\star\tilde{\Gamma}^\star_0 = O(n)\ a.s \tag{54}$$

where $\tilde{\Gamma}^\star_0 = E^\star(\tilde{u}^\star_t)^2$. But the terms in equation (52) are $\frac{1}{n}$ times the square root of (53) and (54). Hence, equation (52) follows. Now, to prove equation (23), recall that $w^\star_t = \sum_{k=1}^{t}\varepsilon^\star_k$ and, from the Beveridge-Nelson decomposition of $u^\star_t$:

$$y^{\star 2}_t = \hat{\pi}_n(1)^2 w^{\star 2}_t + (\tilde{u}^\star_0)^2 + (\tilde{u}^\star_t)^2 - 2\tilde{u}^\star_t u^\star_0 + 2\hat{\pi}_n(1)w^\star_t(\tilde{u}^\star_0 - \tilde{u}^\star_t)$$

thus, $(1/n^2)\sum_{t=1}^{n} y^{\star 2}_t$ is equal to

$$\hat{\pi}_n(1)^2\frac{1}{n^2}\sum_{t=1}^{n}w^{\star 2}_t+\frac{1}{n^2}(\tilde{u}^\star_0)^2+\frac{1}{n^2}\sum_{t=1}^{n}(\tilde{u}^\star_t)^2-\frac{2}{n^2}\tilde{u}^\star_0\sum_{t=1}^{n}\tilde{u}^\star_t+2\hat{\pi}_n(1)\tilde{u}^\star_0\frac{1}{n^2}\sum_{t=1}^{n}w^\star_t+2\hat{\pi}_n(1)\frac{1}{n^2}\sum_{t=1}^{n}w^\star_t\tilde{u}^\star_t.$$

By lemma 3, every term but the first is $o(1)$ a.s. The result follows. ■

Proof of Lemma 5.

Using the definition of bootstrap stochastic orders, we must show that:

$$E^\star\left(\left\|\left(\frac{1}{n}\sum_{t=1}^{n}x^\star_{p,t}x^{\star\,\top}_{p,t}\right)^{-1}\right\|\right) = O_p(1) \tag{55}$$

$$E^\star\left(\left\|\sum_{t=1}^{n}x^\star_{p,t}y^\star_{t-1}\right\|\right) = O_p(np^{1/2})\text{ a.s.} \tag{56}$$

$$E^\star\left(\left\|\sum_{t=1}^{n}x^\star_{p,t}\varepsilon^\star_t\right\|\right) = O_p(n^{1/2}p^{1/2})\text{ a.s.} \tag{57}$$

The proofs below rely on the fact that, under the null, the ADF regression is a finite order autoregressive approximation to the bootstrap DGP, which has an AR(∞) form. To prove (55), let us first define the long run covariance of the vector $x^\star_{p,t}$ as $\Omega^\star_{pp} = (\Gamma^\star_{i-j})_{i,j=1}^{p}$. Then, recalling the result of lemma A5, we have that

$$E^\star\left(\left\|\frac{1}{n}\sum_{t=1}^{n}x^\star_{p,t}x^{\star\,\top}_{p,t} - \Omega^\star_{pp}\right\|^2\right) = O_p(n^{-1}p^2)\text{ a.s.} \tag{58}$$

This is because equation (58) is squared with a factor of $1/n$ and the dimension of $x^\star_{p,t}$ is p. Also,

$$\left\|\Omega^{\star\,-1}_{pp}\right\| \le \left[2\pi(\inf_\lambda f^\star(\lambda))\right]^{-1} = O(1)\text{ a.s.} \tag{59}$$

because, by lemma A4, we can apply the result from Berk (1974, equation 2.14). By assumption 1 (b) and the results of lemma 1, we may say that $\sum_{k=0}^{\infty}|\hat{\pi}_{q,k}| <$

∞. Therefore, as argued by Berk (1974, p. 493), the polynomial $1+\sum_{k=1}^{q}\hat{\pi}_{q,k}e^{ik\lambda}$ is continuous and nonzero over λ so that $f^{\star}(\lambda)$ is also continuous and there are constant values F_1 and F_2 such that $0<F_1<f^{\star}(\lambda)<F_2$. This further implies that (Grenander and Szegö 1958, p. 64) $2\pi\mathrm{F}_1\leq\lambda_1<...<\lambda_p\leq 2\pi F_2$ where λ_i, $i=1,...,p$ are the eigenvalues of the covariance matrix of the bootstrap DGP. Equation (59) then follows from the definition of the matrix norm. See Berk (1974, p. 492-93) for details. It then follows that

$$\left|E^{\star}\left(\left\|\frac{1}{n}\sum_{t=1}^{n}x_{p,t}^{\star}x_{p,t}^{\star}{}^{\top}\right\|\right)-\left\|\Omega_{pp}^{\star}{}^{-1}\right\|\right|\leq E^{\star}\left(\left\|\left(\frac{1}{n}\sum_{t=1}^{n}x_{p,t}^{\star}x_{p,t}^{\star}{}^{\top}\right)^{-1}-\Omega_{pp}^{\star}{}^{-1}\right\|\right)$$

where we used the fact that $E^{\star}\left(\left\|\Omega_{pp}^{\star}\right\|\right)=\left\|\Omega_{pp}^{\star}\right\|$. By equation (58), the right hand side goes to 0 as n increases. Equation (55) then follows from equation (59).

Proof of (56): Our proof is almost identical to the proof of lemma 3.2 in Chang and Park (2002), except that we consider bootstrap quantities. As them, we let $y_t^{\star}=0$ for all $t\leq 0$ and, for $1\leq j\leq p$, we write

$$\sum_{t=1}^{n}y_{t-1}^{\star}u_{t-j}^{\star}=\sum_{t=1}^{n}y_{t-1}^{\star}u_{t}^{\star}+R_n^{\star} \tag{60}$$

where

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

Note that $\sum_{t=1}^{n}x_{p,t}^{\star}y_{t-1}^{\star}$ is a $1\times p$ vector whose j^{th} element is $\sum_{t=1}^{n}y_{t-1}^{\star}u_{t-j}^{\star}$. Therefore, by the definition of the Eucledian norm, equation (56) will be proved if we show that $R_n^{\star}=O_p^{\star}(n)$ uniformly in j for j from 1 to p. We begin by noting that

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

for each j. This allows us to write:

$$R_n^{\star}=\sum_{t=1}^{n}(y_{t-1}^{\star}-y_{t-j-1}^{\star})u_{t-j}^{\star}-\sum_{t=n-j+1}^{n}y_{t-1}^{\star}u_{t}^{\star}.$$

Let us call $R_{1n}^{\star}$ and $R_{2n}^{\star}$ the first and second elements of the right hand side of this last equation. Then, because $y_t^{\star}$ is integrated,

$$\begin{aligned}R_{1n}^{\star}&=\sum_{t=1}^{n}\left(y_{t-1}^{\star}-y_{t-j-1}^{\star}\right)u_{t-j}^{\star}\\&=\sum_{t=1}^{n}\left(\sum_{i=1}^{j}u_{t-i}^{\star}\right)u_{t-j}^{\star}\end{aligned}$$

$$= n\left(\sum_{i=1}^{j}\Gamma^{\star}_{i-j}\right) + \sum_{i=1}^{j}\left[\sum_{t=1}^{n}\left(u^{\star}_{t-i}u^{\star}_{t-j} - \Gamma^{\star}_{i-j}\right)\right]$$

where we have added and subtracted $\Gamma^{\star}_{ij}$. By lemma A3, the first part is $O(n)$ a.s. because $\sum_{-\infty}^{\infty}\Gamma_k = O(1)$. Similarly, we can use lemma A5 to show that the second part is $O^{\star}_p(n^{1/2}p)$ a.s., where the change to $O^{\star}_p$ is because the result of lemma A5 is for the expectation under the bootstrap DGP. The $n^{1/2}$ and p factors appear because lemma A5 considers the square of the present term and j goes from 1 to p. Thus, $R^{\star}_{1n} = O(n) + O^{\star}_p(n^{1/2}p)$ a.s. Let us now consider $R^{\star}_{2n}$:

$$R^{\star}_{2n} = \sum_{t=n-j+1}^{n} y^{\star}_{t-1}u^{\star}_t$$

$$= \sum_{t=n-j+1}^{n}\left(\sum_{i=1}^{t-1}u^{\star}_{t-i}\right)u^{\star}_t$$

$$= \sum_{t=n-j+1}^{n}\sum_{i=1}^{n-j}u^{\star}_t u^{\star}_{t-i} + \sum_{t=n-j+2}^{n}\sum_{i=n-j+1}^{t-1}u^{\star}_t u^{\star}_{t-i}$$

Letting ${R^{\star}_{2n}}^{a}$ and ${R^{\star}_{2n}}^{b}$ denote the first and second part of this last equation, we have

$${R^{\star}_{2n}}^{a} = j\left(\sum_{i=1}^{n-j}\Gamma^{\star}_i\right) + \sum_{t=n-j+1}^{n}\left[\sum_{i=1}^{n-j}(u^{\star}_t u^{\star}_{t-i} - \Gamma^{\star}_i)\right]$$

$$= O(p) + O^{\star}_p(n^{1/2}p) \text{ a.s.}$$

uniformly in j, where the last line comes from lemmas A3 and A5. Similarly, we also have

$${R^{\star}_{2n}}^{b} = (j-1)\sum_{i=n-j+1}^{t-1}\Gamma_i + \sum_{t=n-j+2}^{n}\left[\sum_{i=n-j+1}^{t-1}(u^{\star}_t u^{\star}_{t-i} - \Gamma^{\star}_i)\right]$$

$$= O(p) + O^{\star}_p(p^{3/2}) \text{ a.s.}$$

uniformly in j under lemmas A3 and A5. Hence, $R^{\star}_n$ is $O^{\star}_p(n)$ a.s. uniformly in j. Also, under assumptions 1 and 4, and by lemma 1, $\sum_{t=1}^{n} y^{\star}_{t-1}u^{\star}_t = O^{\star}_p(n)$ a.s. uniformly in j. Therefore, equation (60) is also $O^{\star}_p(n)$ a.s. uniformly in j, $1 \leq j \leq p$. The result follows because the left hand side of (56) is the Euclidian vector norm of p elements that are all $O^{\star}_p(n)$ a.s.

Proof of (57): We begin by noting that for all k such that $1 \leq k \leq p$,

$$E^{\star}\left(\sum_{t=1}^{n}u^{\star}_{t-k}\varepsilon^{\star}_t\right)^2 = n\sigma^2_{\star}\Gamma^{\star}_0$$

which means that
$$E^{\star}\left(\left\|\sum_{t=1}^{n} x_{p,t}^{\star}\varepsilon_{t}^{\star}\right\|^{2}\right) = np\sigma_{\star}^{2}\Gamma_{0}^{\star}.$$

But it has been shown in lemma A2 that $\sigma_{\star}^{2}$ and $\Gamma_{0}^{\star}$ are $O(1)$ a.s. The result is obvious. ■

Proof of Theorem 3.

The theorem follows directly from lemmas 4 and 5. ■

Proof of Corollary 4.

The results are easily obtained by using the Beveridge-Nelson decomposition of the infinite MA form of the ARMASB DGP. ■

Proof of corollary 5.

The proof of lemma 5 can easily be adapted to the ARMASB since lemmas 1, 2 and A1 to A5 have been shown to hold for this case. ■

Proof of Theorem 4

The proof follows directly from corollaries 4 and 5. ■

References

An, H. Z., G. Z. Chan and E. J. Hannan (1982) Autocorrelation, autoregression and autoregressive approximation. *Annals of Statistics* 10, 926-36.

Baxter, G. (1962) An asymptotic result for the finite predictor. *Mathematica Scandinavia* 10, 137-44.

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

Bickel, P. J. and P. Bühlmann (1999) A new mixing notion and functional central limit theorems for a sieve bootstrap in time series. *Bernoulli* 5, 413-46.

Bühlmann, P. (1997) Sieve bootstrap for time series. *Bernoulli* 3, 123-48.

Bühlmann, P. (1998) Sieve bootstrap for smoothing in nonstationarity time series. *Annals of Statistics* 26, 48-83.

Chang, Y. and J. Y. Park (2002) On the asymptotics of ADF tests for unit roots. *Econometric Reviews* 21, 431-47.

Chang, Y. and J. Y. Park (2003) A sieve bootstrap for the test of a unit root. *Journal of Time Series Analysis* 24, 379-400.

Dickey, D. A. and W. A. Fuller (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427-31.

Dickey, D. A. and W. A. Fuller (1981) Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49, 1057-72.

Galbraith, J. W. and V. Zinde-Walsh (1994) A simple noniterative estimator for moving average models. *Biometrika* 81, 143-55.

Galbraith, J. W. and V. Zinde-Walsh (1997) On some simple, autoregression-based estimation and identification techniques for ARMA models. *Biometrika* 84, 685-96.

Grenander, U and G. Szegö (1958) *Toeplitz forms and their applications.* Berkeley, University of California Press.

Hannan, E. J. and L. Kavalieris (1986) Regression, autoregression models. *Journal of Time Series Analysis* 7, 27-49.

Kreiss, J. P. (1992) Bootstrap procedures for AR(∞) processes. In K. H. Jöckel, G. Rothe and W. Senders (eds), *Bootstrapping and Related Techniques, Lecture Notes in Economics and Mathematical Systems 376*, Heidelberg, Springer.

Móricz, F. (1976) Moment inequalities and the strong law of large numbers. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebeite* 35, 298-314.

Ng. S. and P. Perron (1995) Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association* 90, 268-81.

Ng. S. and P. Perron (2001) Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519-54.

Palm, F. C., S. Smeekes and J. P. Urbain (2006) Bootstrap unit root tests: comparison and extensions. working paper, Universiteit Maastricht.

Park, J. Y. (2002) An invariance principle for sieve bootstrap in time series. *Econometric Theory* 18, 469-90.

Park, J. Y. (2003) Bootstrap Unit root tests. *Econometrica*, 71, 1845-95.

Parker, C., Paparoditis, E. and D. N. Politis (in press) Unit root testing via the stationary bootstrap. *Journal of Econometrics*.

Paparoditis, E. and D. N. Politis (2003) Residual based block bootstrap for unit root testing. *Econometrica* 71, 813-55.

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

Psaradakis, Z. (2001) Bootstrap tests for a autoregressive unit root in the presence of weakly dependent errors. *Journal of Time Series Analysis* 22, 577-94.

Sakhanenko, A. I. (1980) On unimprovable estimates of the rate of convergence in invariance principle. *Nonparametric Statistical Inference, Colloquia Mathematica Societatis János Bolyai* 32, 779-83.

Swensen, A. R. (2003) Bootstrapping unit root tests for integrated processes. *Journal of Time Series Analysis* 24, 99-126.

Figures

Figure 1. Rejection probability at nominal level 5%, MA(1), N=250.

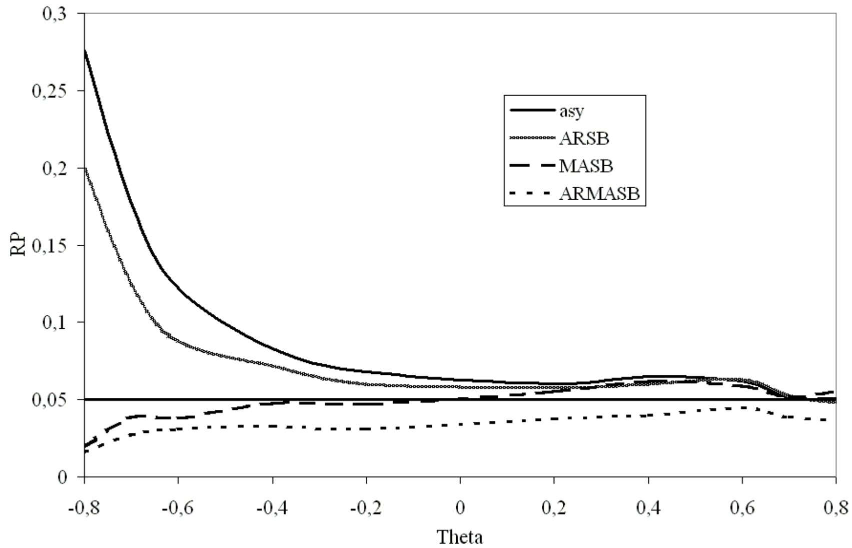

Figure 2. Rejection probability at nominal level 5%, MA(33), N=250.

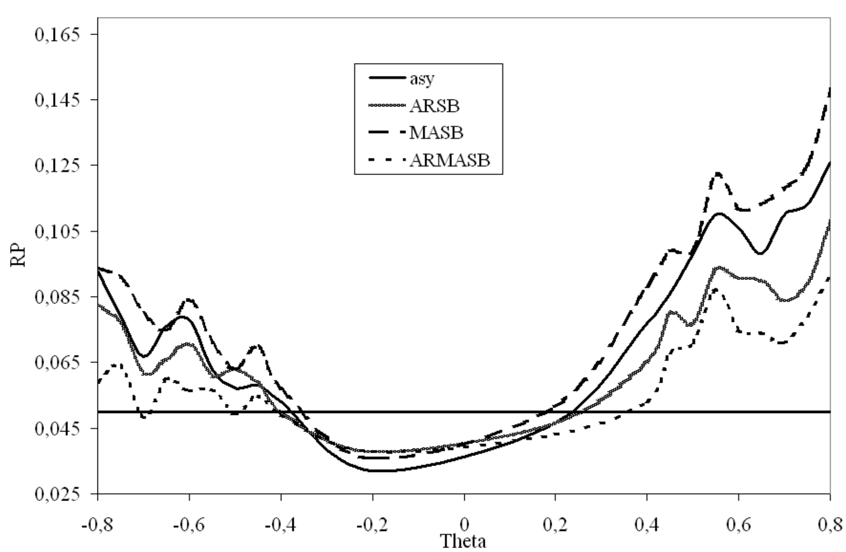

Figure 3. Power at nominal level 5%, MA(1), N=250.