UNIFORM INFERENCE IN AUTOREGRESSIVE MODELS

By Anna Mikusheva [1]

Abstract

The purpose of this paper is to provide theoretical justification for some existing methods of constructing confidence intervals for the sum of coefficients in autoregressive models. We show that the methods of Stock (1991), Andrews (1993), and Hansen (1999) provide asymptotically valid confidence intervals, whereas the subsampling method of Romano and Wolf (2001) does not. In addition, we generalize the three valid methods to a larger class of statistics. We also clarify the difference between uniform and point-wise asymptotic approximations, and show that a point-wise convergence of coverage probabilities for all values of the parameter does not guarantee the validity of the confidence set.

`Key Words`: autoregressive process, confidence set, local to unity asymptotics, uniform convergence

1 Introduction.

Over the past fifteen years, there has been a considerable amount of theoretical and applied work on the problem of constructing a confidence interval for the autoregressive coefficient in an autoregression of order one (AR(1)), or more generally for the sum of the coefficients in an autoregression of order p (AR(p)). From an empirical perspective, the problem is important because the sum of the AR coefficients measures the persistence of a shock to a process. Some recent empirical papers involving confidence intervals for the autoregressive coefficient ρ (or sum of the coefficients) include Murray and Papell (2002), Imbs, Mumtaz, Ravn, and Rey (2005) (exchange

[1] I would like to thank Graham Elliott, Bruce Hansen, Rustam Ibragimov, and Allie Schwartz for comments and suggestions. Suggestions of four anonymous referees and a coeditor greatly improved the paper. I am grateful to Marcelo Moreira for numerous discussions, his enthusiasm and support. I am deeply indebted to Jim Stock for posing the question, for his help, support and encouragement. I would like to thank Aleksander Bulinski for introducing me to the theory of stochastic processes and Skorokhod's embedding. I thank Donald Andrews and Patrik Guggenberger for finding some technical difficulties in an earlier version of the paper. All remaining errors are mine.

rate dynamics), Rapach and Wohar (2004) (real interest rates), and O'Reilly and Whelan (2005) (inflation). From a theoretical perspective, the problem is of interest because the finite-sample distribution of the OLS estimator of the AR(1) coefficient is biased, its limiting distribution (and its rate) changes near $\rho = 1$, and there is no known pivotal statistic.

Several methods have been proposed for constructing confidence intervals for ρ. Stock (1991) proposed inverting the Dickey-Fuller (1979) t-statistic when ρ is in a $1/T$ neighborhood of one (local to unity). Andrews (1993) suggested inverting the finite sample distribution of the OLS estimator of ρ under the assumption of normality. Hansen (1999) introduced the "grid bootstrap" method, which is conceptually similar to Andrews' method except that Hansen inverted the t-statistic and used the bootstrap to approximate the exact distribution on a grid of values of ρ. Romano and Wolf (2001) proposed a subsampling method, in which the distribution of the statistics in subsamples is used to approximate the sampling distribution. Yet, as these authors recognize, all of these methods have limitations, and none of the existing proofs actually prove that the confidence intervals are asymptotically correct under standard weak assumptions on the errors.[2]

In this paper we focus on confidence sets that in large samples have correct coverage *uniformly* over the parameter space. We refer to these as "valid" confidence intervals. For a confidence set C to be asymptotically valid, the following condition must hold: $\lim_{T\to\infty} \inf_\rho P_\rho(\rho \in C) \geq 1 - \alpha$, where the infimum is taken over the parameter space (for example, $|\rho| \leq 1$). That is, the notion of "validity" in this paper is closely related to the concept of "global uniform validity".

The distinction between uniform validity and point-wise validity(that is, $\inf_\rho \lim_{T\to\infty} P_\rho(\rho \in C) \geq 1 - \alpha$) is not of practical importance in many econometric applications, but

[2]Through simulations, Andrews (1991) showed that the method seems to be robust to non-normal errors, but did not prove that it can be used in a general AR model. Stock (1991) proved the validity of his method in the local-to-unity neighborhood, but there is no proof that this method can be used when ρ is fixed and less than one. Hansen (1999) proved that his grid bootstrap provides point-wise asymptotically correct approximations under both classical and local to unity asymptotics. Romano and Wolf (2001) prove point-wise validity for $|\rho| < 1$ and for $\rho = 1$.

it is here. For example, consider the pretest confidence set, constructed as the OLS estimator ± 1.96 standard errors if the Dickey-Fuller t-statistic is less than -ln T, and which equals one otherwise. This set satisfies the point-wise convergence criterion but not the uniform criterion because the procedure places point mass on one for all values of ρ in a $1/T$ neighborhood of one. For this set, the asymptotic coverage rate is actually zero, that is, there are (sequences of) parameter values for which the probability of being in the set tends to zero, no matter how large the sample size. Lack of uniformity is one way to understand the poor performance of standard confidence intervals in the AR(1) model discussed by Nankervis and Savin (1985, 1988) and Rayner (1990).

The purpose of this paper is to prove the uniform validity or invalidity of a variety of methods for the construction of confidence intervals for ρ. The proofs use recently developed tools involving the strong invariance principal and stochastic process theory. We show that the methods of Stock (1991), Andrews (1993), and Hansen (1999), when based on inverting the t-statistic, provide asymptotically valid confidence intervals in the uniform sense stated above, but the subsampling method of Romano and Wolf (2001) does not(even though it possesses a point-wise validity). In addition to these main results, we generalize the three valid methods to a larger class of statistics.

The paper proceeds as follows. The next section introduces the concept of uniform asymptotic approximation and relates it to the construction of asymptotically correct confidence sets. It also sets up the theoretical framework for proving validity of the three methods. Sections 3-6 consider the AR (1) model with an intercept and/or a linear time trend. Section 3 provides large sample justification for the method proposed by Andrews (1993). Section 4 proves that the local to unity asymptotic approach (Stock (1991)) provides a uniform approximation even for values of the AR coefficient far from the unit root. Section 5 gives a theoretical justification of Hansen's (1999) grid bootstrap. In section 6 we show why the Romano and Wolf (2001) subsampling method should not be used for making inferences in an AR model. Section 7 extends the results to AR(p) processes with at most one root close to the unit circle. The Appendix contains proofs of results from Sections 2 - 6.

Since the proofs are long and technically involved, we place some details of the proofs and the proofs of the results from Section 7 in the Supplementary Appendix, which can be found on the author's web-site.[3] The Supplementary Appendix also contains a simulation study assessing the finite sample properties of the discussed procedures.

2 Uniform approximation.

Assume that we have a sample $Y = (y_1, ..., y_T)$ of size T from an AR(1) process with an intercept:

$$y_j = c + x_j; \quad x_j = \rho x_{j-1} + \varepsilon_j, \ j = 1, ..., T, \quad x_0 = 0. \tag{1}$$

The autoregressive coefficient ρ can take on any values in the open interval $\Theta = (-1, 1)$. Model (1) is often used in practice to describe the behavior of inflation or the logarithm of exchange rate. In Section 7 we show that the results can be extended to more general autoregressive processes. We make the following assumptions about the error terms:

Assumptions A. Let $(\varepsilon_j, \mathcal{F}_j)$ be a martingale - difference sequence with $E(\varepsilon_j^2|\mathcal{F}_{j-1}) = 1$ and $\sup_j E(|\varepsilon_j|^r|\mathcal{F}_{j-1}) < \infty$ a.s. for some $2 < r \leq 4$.

We are interested in constructing a confidence set for the parameter ρ. Below is the classical definition of a confidence set (Lehmann (1997), p.90).

Definition A subset $C(Y)$ of the parameter space Θ is said to be a *confidence set* at a confidence level $1 - \alpha$ if $\inf_{\rho \in \Theta} P_\rho\{\rho \in C(Y)\} \geq 1 - \alpha$.

Definition A subset $C(Y)$ of the parameter space Θ is said to be an *asymptotic confidence set* at a confidence level $1 - \alpha$ (or is said to have a uniform asymptotic coverage probability $1 - \alpha$) if

$$\liminf_{T \to \infty} \inf_{\rho \in \Theta} P_\rho\{\rho \in C(Y)\} \geq 1 - \alpha. \tag{2}$$

The requirement of uniform convergence (2) is much stronger than a requirement

[3] http://www.people.fas.harvard.edu/ mikouch/paper-1/appendix.pdf

of point-wise convergence of coverage probabilities

$$\lim_{T\to\infty} P_\rho\{\rho \in C(Y)\} \geq 1-\alpha \ \text{ for every } \ \rho \in \Theta. \tag{3}$$

The convergence (3) says that for every value of the parameter space and for any given accuracy we can find a large enough sample size providing the required accuracy of coverage at this value. However, convergence at some values of the parameter can be much slower than at others. Condition (3) does not guarantee that there is a sample size providing the required accuracy for *all* values of the parameter. That is, even for a huge sample size we might find a part of the parameter space where the required accuracy has not been achieved. Since a priori we cannot guarantee that our parameter does not belong to this part of the parameter space, we are always at risk of having poor coverage probability.

This paper deals with methods based on the classical idea of inverting tests (Lehmann (1997), p.90). Let $A(\rho_0)$ be an acceptance region of an asymptotic level - α test for testing $H_0 : \rho = \rho_0$. A set $C(Y)$ is constructed as a set of parameter values for which the corresponding simple hypothesis is accepted $C(Y) = \{\rho : Y \in A(\rho)\}$.

Let the testing procedure for a test of the hypothesis $H_0 : \rho = \rho_0$ be based on a test statistic $\varphi(Y, T, \rho_0)$ and critical values $c_1(T, \rho_0)$ and $c_2(T, \rho_0)$. A set $C(Y)$ is defined as

$$C(Y) = \{\rho \in \Theta : c_1(T, \rho) \leq \varphi(Y, T, \rho) \leq c_2(T, \rho)\}. \tag{4}$$

We state all results for two - tailed tests, but they are equally as applicable for one - tailed tests.

Let $F_{T,\rho}(x) = P_\rho\{\varphi(Y, T, \rho) \leq x\}$ be a distribution function of the statistic $\varphi(Y, T, \rho)$ given that the true AR parameter is equal to ρ. Let $q^F_\alpha(T, \rho)$ denote an α - quantile of the distribution $F_{T,\rho}(x)$, that is, $F_{T,\rho}(q^F_\alpha(T, \rho)) = \alpha$. If $c_1(T, \rho) = q^F_{\alpha/2}(T, \rho)$ and $c_2(T, \rho) = q^F_{1-\alpha/2}(T, \rho)$, then a set $C(Y)$ defined by (4) is a confidence set at the confidence level $1 - \alpha$. In practice, the finite sample distribution $F_{T,\rho}(x)$ is usually unknown. However, if there is a family of distributions that provides a *uniform* asymptotic approximation of $F_{T,\rho}(x)$, then it can be used to construct an asymptotic confidence set.

Lemma 1 *Let $G_{T,\rho}(x)$ be a family of distribution functions uniformly approximating the family of distributions $F_{T,\rho}(x)$ as T increases:*

$$\lim_{T\to\infty}\sup_{\rho\in\Theta}\sup_{x}|F_{T,\rho}(x)-G_{T,\rho}(x)|=0.$$

Suppose, that a set $C(Y)$ is defined by (4) with $c_1(T,\rho)=q^G_{\alpha/2}(T,\rho)$ and $c_2(T,\rho)=q^G_{1-\alpha/2}(T,\rho)$, where $q^G_\alpha(T,\rho)$ is the α - quantile of the distribution $G_{T,\rho}(x)$. Then $C(Y)$ is an asymptotic confidence set at the confidence level $1-\alpha$.

Remark 1 *Having a uniformly approximating family of distributions is a sufficient, but not necessary condition for constructing an asymptotic confidence set. The real line $\mathbb{R}$ is an asymptotic confidence set at any level. However, this set is useless from the practical point of view, since it has zero power. The confidence set constructed in Lemma 1 using a uniform approximation is not conservative. In particular,*

$$\lim_{T\to\infty}\inf_{\rho\in\Theta}P_\rho\{\rho\in C(Y)\}=\lim_{T\to\infty}\sup_{\rho\in\Theta}P_\rho\{\rho\in C(Y)\}=1-\alpha.$$

The main goal of this paper is to prove that the three methods widely used in practice: Andrews' parametric grid bootstrap, Stock's method based on the local to unity asymptotic approach and Hansen's grid bootstrap provide asymptotic confidence sets via constructing uniformly approximating families of distributions. The rest of the section describes a joint theoretical framework for the proofs of all three methods.

2.1 Class of Test Statistics

Let y^μ_j be the demeaned process of y_j, that is, $y^\mu_j=y_j-\frac{1}{T}\sum_{i=1}^T y_{i-1}$. We consider a wide class of test statistics based on a pair of statistics

$$(S(T,\rho),R(T,\rho))=\left(\frac{1}{\sqrt{g(T,\rho)}}\sum_{j=1}^T y^\mu_{j-1}(y_j-\rho y_{j-1}),\frac{1}{g(T,\rho)}\sum_{j=1}^T(y^\mu_{j-1})^2\right),$$

where $g(T,\rho)$ is a normalization function. We define $g(T,\rho)=E_\rho\left(\sum_{j=1}^T(y^\mu_{j-1})^2\right)$. We should note that statistics (S,R) are invariant with respect to values of c.

Most of the results of the paper also hold for slightly explosive processes. Let us introduce a sequence of sets $\Theta_T = [-1 - \theta/T, 1 + \theta/T]$ of possible values of the AR coefficient when a sample size equals T, here $\theta > 0$.

Definition Let H be the class of functions $\phi(s, r, T, \rho) : \mathbb{R} \times \mathbb{R}_+ \times \mathbb{N} \times \Theta_1 \to \mathbb{R}$ satisfying two conditions:
1) for every $C > 0$ there exist constants M_C and T_1 such that for all $s, s_1 \in \mathbb{R}, r > C, r_1 > C, T > T_1, \rho \in \Theta_T$ we have $|\phi(s, r, T, \rho) - \phi(s_1, r_1, T, \rho)| < M_C(|s - s_1| + |r - r_1|)$;
2) for every $0 < C_1 < C_2 < \infty$ there exists a constant $A > 0$ such that $\frac{\partial \phi(s,r,T,\rho)}{\partial s} > A$ for all T, $\rho \in \Theta_T$ and $C_1 < r < C_2$.

Definition *The class $\mathcal{H}$ of test statistics* under consideration is given by the following set $\mathcal{H} = \{\varphi(Y, T, \rho) = \phi(S(T, \rho), R(T, \rho), T, \rho) : \phi(s, r, T, \rho) \in H\}$.

The class $\mathcal{H}$ is quite wide. For instance, it includes the conventional t-statistic $t = \frac{S}{\sqrt{R}}$, and an appropriately normalized OLS estimate of the autoregressive coefficient $\sqrt{g(T, \rho)}(\widehat{\rho}_{OLS} - \rho) = \frac{S}{R}$.

Selecting the proper test statistic is a difficult task. More powerful testing procedures tend to produce more accurate confidence sets. It is well-known that even in model (1) without an intercept and with normal errors, the uniformly most powerful test for the simple hypothesis $H_0 : \rho = \rho_0$ does not exist (Dufour and King (1991), Elliott, Rothenberg and Stock (1996)). The class $\mathcal{H}$ contains all test statistics used to create the power envelope considered in Elliott, Rothenberg and Stock (1996). Our class of test statistics allows different test statistics for testing different values of ρ. The idea goes back to Elliott and Stock (2001), who suggested inverting a sequence of point optimal tests. They showed that the confidence intervals constructed from inverting a sequence of point optimal tests have quite similar power properties to inverting near optimal tests for a unit root.

2.2 Stationary and near unity regions

Let a set $C(Y)$ be defined by (4), where the test statistic is given by $\varphi(Y, T, \rho) = \phi(S, R, T, \rho)$ and the critical value functions $c_1(T, \rho)$ and $c_2(T, \rho)$ are calculated using one of the three methods mentioned above. In all three methods $c_1(T, \rho)$ and $c_2(T, \rho)$

are quantiles of the distribution of an approximating statistic $\phi(S_1(T,\rho), R_1(T,\rho), T, \rho)$. In Andrews' method the approximating pair of statistics $(S_1(T,\rho), R_1(T,\rho)) = (S^N, R^N)$ is calculated using an AR(1) process with normal errors. In Stock's method we calculate the limiting distribution (S^c, R^c) of the pair $(S(T,\rho), R(T,\rho))$ when the limit is taken along a sequence of models with $\rho_T = \exp\{c/T\}$. The we use $(S_1(T,\rho), R_1(T,\rho)) = (S^{c(T,\rho)}, R^{c(T,\rho)})$ as the approximating pair. Hansen's grid bootstrap approximates the distribution of (S, R) by the distribution of a pair of bootstrapped statistics (S^*, R^*).

We want to show that the distribution of the statistic $\varphi_1(T,\rho) = \phi(S_1, R_1, T, \rho)$ approximates the unknown finite sample distribution of the statistic $\varphi(T,\rho) = \phi(S, R, T, \rho)$ uniformly over ρ:

$$\lim_{T\to\infty} \sup_{\rho\in\Theta_T} \sup_x |P\{\varphi(T,\rho) < x\} - P\{\varphi_1(T,\rho) < x\}| = 0. \tag{5}$$

There are two different asymptotic approaches developed for autoregressive processes. These approaches describe strikingly different asymptotic behavior depending on how close the parameter ρ is to unit root.

The classical approach is based on the Central Limit Theorem and the Law of Large Numbers. If $|\rho| < 1$ is fixed then

$$\sqrt{\frac{1-\rho^2}{T}} \sum_{j=1}^{T} x_{j-1}\varepsilon_j \Rightarrow N(0,1) \text{ and } \frac{1-\rho^2}{T} \sum_{j=1}^{T} x_{j-1}^2 \to^p 1 \quad \text{as } T \to \infty. \tag{6}$$

Park (2003) and Giraitis and Phillips (2006) generalized this result for sequences of processes for which the autoregressive coefficient ρ_T converges to the unit root as the sample size increases with speed slower than $1/T$.

The second asymptotic approach, local to unity asymptotics, was developed by Bobkoski (1983), Cavanagh (1985), Chan and Wei (1987), and Phillips (1987). If the autoregressive coefficient is local to unity , i.e. is defined by a sequence $\rho_T = 1+c/T$ for some fixed $c < 0$, then as the sample size increases we have the following convergence:

$$\left(\frac{1}{T}\sum_{j=1}^{T} x_{j-1}\varepsilon_j, \frac{1}{T^2}\sum_{j=1}^{T} x_{j-1}^2\right) \Rightarrow \left(\int_0^1 J_c(x)dw(x), \int_0^1 J_c^2(x)dx\right), \tag{7}$$

where the process J_c is an Ornstein - Uhlenbeck process defined by $J_c(x) = \int_0^x e^{(x-y)c}dw(y)$, and $w(y)$ is a Brownian motion.

Along with the different behaviors of the autoregressive coefficient, both asymptotics require different normalizations of the sums. The classical asymptotic approach employs the normalization $(\frac{1}{\sqrt{T}}, \frac{1}{T})$, whereas the local to unity asymptotic approach uses $(\frac{1}{T}, \frac{1}{T^2})$. The proposed normalization $\left(\frac{1}{\sqrt{g(T,\rho)}}, \frac{1}{g(T,\rho)}\right)$ joins the two in the same framework. For any fixed $|\rho| < 1$, our normalization is asymptotically proportional to $(\frac{1}{\sqrt{T}}, \frac{1}{T})$, and for $\rho_T = 1 + c/T$ it is asymptotically proportional to $(\frac{1}{T}, \frac{1}{T^2})$. For fixed $|\rho| < 1$ the statistic $S(T, \rho)$ has an asymptotically standard normal distribution, and $R(T, \rho)$ converges to 1 in probability. Along the sequence $\rho_T = 1 + c/T$ the statistics $(S(T, \rho_T), R(T, \rho_T))$ converge weakly to a pair of non-normal distributions.

In order to receive the uniform approximation (5) we divide the parameter space Θ_T into two overlapping regions, the "stationary" and the "near unity" regions. The stationary region is separated from the unit root by a neighborhood contracting at a speed slower than $1/T$. The near unity area shrinks toward the unit root at an even slower speed.

The pair (S, R) and an approximating pair (S_1, R_1) follow the classical asymptotic behavior in the stationary region. Namely, statistics S and S_1 weakly converge to the standard normal distribution, whereas statistics R and R_1 converge in probability to one. The convergence in both cases is uniform over the stationary region.

Approximating in the near unity region is a more delicate task. We will be able to construct pairs (S, R) and (S_1, R_1) on a common probability space in such a way that the distance between them converges to zero in probability uniformly over the near unity region.

The general framework of the proofs of uniformity for the three methods is formalized in the lemma below:

Lemma 2 *Let $(S(T, \rho), R(T, \rho))$ and $(S_1(T, \rho), R_1(T, \rho))$ be two pairs of random functions. Assume that there exists a sequence of overlapping sets $\mathcal{A}_T$ and $\mathcal{B}_T$ such that $\mathcal{A}_T \bigcup \mathcal{B}_T = \Theta_T$. Let the following conditions be satisfied:*

1. We can define variables $(S(T, \rho), R(T, \rho))$ and $(S_1(T, \rho), R_1(T, \rho))$ on a common *probability space in such a way that for every* $\varepsilon > 0$

$$\lim_{T\to\infty} \sup_{\rho\in\mathcal{A}_T} P\left\{|S(T,\rho) - S_1(T,\rho)| + |R(T,\rho) - R_1(T,\rho)| > \varepsilon\right\} = 0;$$

2. There exists a continuous distribution function $F(x)$ *that does not depend on either* T *or* ρ*, such that* S *and* S_1 *both converge in distribution to* $F(x)$ *uniformly over* $\mathcal{B}_T$*:*

$$\lim_{T\to\infty} \sup_{\rho\in\mathcal{B}_T} \sup_x |P\{S(T,\rho) < x\} - F(x)| = \lim_{T\to\infty} \sup_{\rho\in\mathcal{B}_T} \sup_x |P\{S_1(T,\rho) < x\} - F(x)| = 0.$$

3. As the sample size increases, R *and* R_1 *both converge uniformly over* $\mathcal{B}_T$ *to the same constant* K *that does not depend on* ρ *:*

$$\lim_{T\to\infty} \sup_{\rho\in\mathcal{B}_T} P\{|R(T,\rho) - K| > \varepsilon\} = \lim_{T\to\infty} \sup_{\rho\in\mathcal{B}_T} P\{|R_1(T,\rho) - K| > \varepsilon\} = 0 \quad \forall\varepsilon > 0.$$

4. For every $\varepsilon > 0$ *there exists* $C > 0$ *such that* $\sup_T \sup_{\rho\in\Theta_T} P\{R_1(T,\rho) < C\} < \varepsilon$*; that is,* R_1 *is separated from zero uniformly over* Θ_T*. We also assume that* $ER_1 = ER = K$*.*

5. The pair of variables (S_1, R_1) *possesses a continuous distribution uniformly over* ρ *in the following way: for every* $\varepsilon > 0$ *there exists a constant* $M > 0$ *such that for all* $\delta_1 < \varepsilon, \delta_2 < \varepsilon$*,* $|b - K| > 2\varepsilon$ *and all* $\rho \in \Theta_T$ *and* T *we have:*

$$P_\rho\{(S_1(T,\rho), R_1(T,\rho)) \in [a - \delta_1, a + \delta_1] \times [b - \delta_2, b + \delta_2]\} \leq M\delta_1\delta_2;$$

$$P_\rho\{S_1(T,\rho) \in [a - \delta_1, a + \delta_1]\} \leq M\delta_1.$$

Then for every statistic $\varphi(T,\rho) = \phi(S,R,T,\rho) \in \mathcal{H}$ *its distribution is uniformly approximated by the distribution of* $\varphi_1(T,\rho) = \phi(S_1,R_1,T,\rho)$*. Thus, convergence (5) holds. If* $c_1(T,\rho)$ *and* $c_2(T,\rho)$ *are the quantiles of the distribution of* $\varphi_1(T,\rho)$*, the set* $C(Y)$ *defined by (4) is an asymptotic confidence set.*

The approximation in the near unity region is stated in Condition 1. The asymptotic behavior of the pairs (S,R) and (S_1,R_1) in the stationary region is described in Conditions 2 and 3. The statistic R is allowed to appear in the denominator of a test statistic. For example, the t-statistic can be written as $t = \frac{S}{\sqrt{R}}$. Condition 4 is imposed in order to guarantee that $\frac{1}{R}$ is uniformly bounded. Condition 5 is a technical one. It requires the distribution of S_1 to be uniformly continuous and the joint distribution of (S_1,R_1) to be uniformly continuous in the area where R_1 is separated from its stationary limit K.

2.3 Estimation of variance.

Now we relax the assumption of the previous sections that the conditional variance of errors is known.

Let $\widetilde{Y} = (\widetilde{y}_1, ..., \widetilde{y}_T)$ be a sample from an AR(1) process defined by an equation

$$\widetilde{y}_j = c + \widetilde{x}_j; \quad \widetilde{x}_j = \rho \widetilde{x}_{j-1} + \widetilde{\varepsilon}_j, \; j = 0, ..., T, \quad \widetilde{x}_0 = 0. \tag{8}$$

Error terms satisfy a set of Assumptions A1 stated below.

Assumptions A1. Let$(\widetilde{\varepsilon}_j, \mathcal{F}_j)$ be a martingale difference sequence with $E(\widetilde{\varepsilon}_j^2 | \mathcal{F}_{j-1}) = \sigma^2$ and $\sup_j E(|\widetilde{\varepsilon}_j|^r | \mathcal{F}_{j-1}) < \infty$ a.s. for some $2 < r \leq 4$.

Note that if the variance of error terms σ^2 is known, then the process $y_j = \widetilde{y}_j / \sigma$ is a process described by (1) with errors satisfying the set of Assumptions A.

Let $\widehat{e}_j = \widetilde{y}_j^{\mu} - \widehat{\rho}_{OLS} \widetilde{y}_{j-1}^{\mu}$ be the OLS residuals. Let us define an estimator of σ^2 to be a sample variance of OLS residuals: $\widehat{\sigma}^2 = \frac{1}{T} \sum_{j=1}^{T} \widehat{e}_j^2$. Despite of the fact that the estimator $\widehat{\rho}_{OLS}$ of the AR coefficient is biased toward zero, the estimator $\widehat{\sigma}^2$ of error term variance is uniformly consistent.

Let us define statistics $(\widetilde{S}, \widetilde{R})$ in the following way

$$(\widetilde{S}, \widetilde{R}) = (\frac{1}{\sqrt{g(T,\rho)} \widehat{\sigma}^2} \sum_{j=1}^{T} \widetilde{y}_{j-1}^{\mu} \widetilde{\varepsilon}_j, \frac{1}{g(T,\rho) \widehat{\sigma}^2} \sum_{j=1}^{T} (\widetilde{y}_{j-1}^{\mu})^2).$$

Lemma 3 *Let us consider a model (8) with error terms satisfying the set of Assumptions A1, then* $\lim_{T\to\infty} \sup_{\rho \in \Theta_T} P\left\{ \left| \frac{\widehat{\sigma}^2}{\sigma^2} - 1 \right| > \varepsilon \right\} = 0$ *for every* $\varepsilon > 0$.

Any statistic $\varphi(\widetilde{Y}, T, \rho) = \phi(\widetilde{S}, \widetilde{R}, T, \rho)$ *, where* ϕ *belongs to the class* H*, is uniformly approximated by the corresponding statistic* $\phi(S, R, T, \rho)$*, where the pair* (S, R) *is defined for the process* $y_j = \widetilde{y}_j / \sigma$ *with the unit variance of error terms.*

The proof of Lemma 3 is put to the Supplementary Appendix.

3 Validity of Andrews' method.

This section proves the validity of the method proposed by Andrews (1993). Let us consider an AR(1) model with normal errors:

$$z_j = \rho z_{j-1} + e_j, \quad e_j \sim \; iid \; N(0,1), \quad z_0 = 0. \tag{9}$$

The finite sample distribution of the pair of statistics

$$\left(\frac{1}{\sqrt{g(T,\rho)}}\sum_{j=1}^{T} z_{j-1}^{\mu}e_j, \frac{1}{g(T,\rho)}\sum_{j=1}^{T}(z_{j-1}^{\mu})^2\right) = (S^N, R^N), \tag{10}$$

is fully defined for every T and ρ and can be calculated by numerical integration or by simulations. The distribution of the statistic $\varphi_1 = \phi(S^N, R^N, T, \rho)$ is also fully defined and can be simulated.

Assume that we have a sample $Y = (y_1, ..., y_T)$ from process (1), with independent standard normal error terms ε_t. Then set $C(Y)$, defined by equation (4) where $c_1(T,\rho)$ and $c_2(T,\rho)$ equal to the $\alpha/2$ and $1-\alpha/2$ quantiles of the finite sample distribution of the statistic φ_1, is a confidence set for the parameter ρ at confidence level $1-\alpha$. Andrews (1993) proposed the procedure described above for the test statistic equal to the OLS estimator of ρ, but the procedure can be generalized for any $\varphi \in \mathcal{H}$.

The described procedure is exact only for the AR(1) model with *normal* errors. Andrews (1993) performed simulations showing that the method is robust to non-normal errors. We prove that the method produces asymptotically uniform confidence sets if applied to model (1) without normality assumptions.

Theorem 1 *Let Y be a sample from an AR(1) process with an intercept defined by (1) with error terms satisfying the set of Assumptions A. We consider a test statistic $\varphi(Y,T,\rho) = \phi(S,R,T,\rho)$ belonging to the class $\mathcal{H}$. Let $C(Y)$ be a set defined by equation (4) with $c_1(T,\rho)$ and $c_2(T,\rho)$ being the $\alpha/2$ and $1-\alpha/2$ quantiles of the finite sample distribution of the statistic $\varphi_1 = \phi(S^N, R^N, T, \rho)$, where (S^N, R^N) are statistics defined by (10) for model (9) with normal errors. Then $C(Y)$ has a uniform asymptotic coverage probability equal to $1-\alpha$.*

Remark 2 *The statement of Theorem 1 can be extended to an AR(1) process with a linear trend, $y_j = a + bj + x_j$, where x_j is defined in (1). Let y_j^{τ} be the detrended process of y_j, that is, $y_j^{\tau} = y_j - \overline{y} - \frac{\sum_{i=1}^{T}(y_i-\overline{y})i}{\sum_{i=1}^{T}(i-\frac{T+1}{2})^2}(j - \frac{T+1}{2})$. We consider a pair of statistics*

$$(S^{\tau}(T,\rho), R^{\tau}(T,\rho)) = \left(\frac{1}{\sqrt{g^{\tau}(T,\rho)}}\sum_{j=1}^{T} y_{j-1}^{\tau}(y_j - \rho y_{j-1}), \frac{1}{g^{\tau}(T,\rho)}\sum_{j=1}^{T}(y_{j-1}^{\tau})^2\right),$$

where $g^{\tau}(T,\rho) = E\left(\sum_{j=1}^{T}(y_{j-1}^{\tau})^2\right)$. *Then if Assumptions A are satisfied, the finite sample distribution of the statistic* $\varphi = \phi(S^{\tau}, R^{\tau}, T, \rho)$ *is uniformly approximated by the finite sample distribution of the statistic* $\varphi_1 = \phi(S^{\tau,N}, R^{\tau,N}, T, \rho)$. *Here* $(S^{\tau,N}, R^{\tau,N})$ *is a pair of corresponding detrended statistics in a model with normal errors.*

The proof of Theorem 1 follows the plan proposed in Lemma 2. Let

$$\mathcal{A}_T = \{\rho \in \Theta_T : |1-\rho|T^{\alpha} < 1 \quad \text{or } |1+\rho|T^{\alpha} < 1\} \tag{11}$$

for some $0 < \alpha < 1$ be a near unity region. Let the stationary region be defined by the set

$$\mathcal{B}_T = \{\rho \in \Theta_T : -\rho_T \leq \rho \leq \rho_T\}\,, \quad \text{where} \;\; \rho_T = 1 - \frac{\log(T)}{T}. \tag{12}$$

The sets $\mathcal{A}_T$ and $\mathcal{B}_T$ are overlapping and cover the whole Θ_T.

Giraitis and Phillips (2006) showed that the convergence in (6) holds uniformly over $\mathcal{B}_T$. Conditions 2 and 3 of Lemma 2 are direct corollaries from Lemmas 2.1 and 2.2 in Giraitis and Phillips (2006). The fact that the statistic R^N is uniformly separated from zero (Condition 4) follows from Theorem 2 in Székely and Bakirov (2003).

Our main efforts are devoted to checking Condition 1. Our proof uses the Strong Invariance Principle. We define statistics (S, R) and (S^N, R^N) on a common probability space in such a way that the distance between them converges to zero in probability uniformly over set $\mathcal{A}_T$.

Let $(\varepsilon_j, \mathcal{F}_j)$ be a martingale difference sequence of error terms satisfying the set of Assumptions A. We consider partial sums $S_j = \sum_{i=1}^{j}\varepsilon_i$ and the normalized partial sum process $\eta_T(t) = \frac{1}{\sqrt{T}}S_{[tT]}$. Using Skorohod's embedding scheme we can enlarge the initial probability space and construct a sequence of Brownian motions w_T on it in such a way that for every $\varepsilon > 0$ we have

$$\sup_{0\leq t\leq 1}|\eta_T(t) - w_T(t)| = o(T^{-1/2+1/r+\varepsilon}) \quad a.s. \tag{13}$$

For more details, please refer to Lemma 2 in the Supplementary Appendix.

We should note that since $r > 2$, the distance between the processes in (13) converges to zero with the speed of T raised to a negative power. The normalized process of partial sums for variables with a finite moment of higher order can be better approximated by a Brownian motion.

Let us define error terms by the following equality $\frac{e_{T,j}}{\sqrt{T}} = w_T(j/T) - w_T((j-1)/T)$. The error terms $e_{T,j}$ are constructed on the same probability space as ε_t, and have the standard normal distribution.

In the following analysis we work only with positive AR coefficients $\rho \in \mathcal{A}_T^+$. The proof for negative AR coefficients is similar. Let us define $z_{T,j}(\rho) = \rho z_{T,j-1}(\rho) + e_{T,j}$. Then for every ρ the distribution of $\{z_j\}_{j=1}^T$ is the same as the distribution of $\{z_{T,j}(\rho)\}_{j=1}^T$. In what follows we ignore the difference between them. The lemma below shows that many statistics for the constructed processes will be uniformly close to one another.

Lemma 4 *For every $\varepsilon > 0$ we have*

a) $\sup_{\rho \in \Theta_T^+} \sup_{j=1,\dots,T} \left| \frac{x_j}{\sqrt{T}} - \frac{z_j}{\sqrt{T}} \right| = o(T^{-1/2+1/r+\varepsilon})$ *a.s.;*

b) $\sup_{\rho \in \Theta_T^+} \sup_{j=1,\dots,T} \left| \frac{x_j}{\sqrt{T}} \right| = O(1)$ *a.s. ;*

c) $\left| \frac{1}{\sqrt{T}} \sum_{j=1}^T \eta_T(j/T)\varepsilon_j - \frac{1}{\sqrt{T}} \sum_{j=1}^T w_T(j/T) e_{T,j} \right| = o(T^{-1/2+1/r+\varepsilon})$ *a.s.;*

d) $\sup_{\rho \in \Theta_T^+} \frac{1}{(1-\rho)T+1} \left| \frac{1}{T} \sum_{j=1}^T x_{j-1}\varepsilon_j - \frac{1}{T} \sum_{j=1}^T z_{j-1} e_{T,j} \right| = o(T^{-1/2+1/r+\varepsilon})$ *a.s.;*

e) $\sup_{\rho \in \Theta_T^+} \left| \frac{1}{T^2} \sum_{j=1}^T x_{j-1}^2 - \frac{1}{T^2} \sum_{j=1}^T z_{j-1}^2 \right| = o(T^{-1/2+1/r+\varepsilon})$ *a.s.;*

f) $\sup_{\rho \in \Theta_T^+} \left| \frac{1}{T^{3/2+k}} \sum_{j=1}^T x_{j-1} j^k - \frac{1}{T^{3/2+k}} \sum_{j=1}^T z_{j-1} j^k \right| = o(T^{-1/2+1/r+\varepsilon})$ *a.s.;*

g) $\sup_{\rho \in \mathcal{A}_T^+} |S(T,\rho) - S^N(T,\rho)| = o(T^{3/2+1/r-2\alpha+\varepsilon})$ *a.s.;*

h) $\sup_{\rho \in \mathcal{A}_T^+} |R(T,\rho) - R^N(T,\rho)| = o(T^{1/2+1/r-\alpha+\varepsilon})$ *a.s.*

Statements g) and h) of Lemma 4 imply the validity of Condition 1 of Lemma 2 for the pairs (S, R) and (S^N, R^N).

We should note that statements a) through f) hold for the whole parameter space, whereas g) and h) are stated for a local neighborhood of $\rho = 1$ only. Parts a) - f) use the local to unity normalization, which is too strong for our purposes. In g) and h) we received the right normalization at the price of a lower convergence speed. Since $r > 2$ there is always some $0 < \alpha < 1$ such that we have T in a negative power on the right-hand side of statements g) and h).

4 Validity of Stock's method

Stock (1991) proposed to construct a confidence set for the largest autoregressive root by using local to unity asymptotic approximation. He used a t-statistic for testing $H_0 : \rho = 1$. However, this test statistic, unlike the t-statistic for testing the true ρ, does not belong to the class $\mathcal{H}$. Hansen (1999) showed in simulations that the version of the procedure proposed by Stock (1991) breaks down for values of ρ far from the unit root. In this section we prove the validity of a *modification* of the method proposed by Stock (1991) for constructing confidence intervals with the help of local to unity asymptotics.

It is well known from Bobkoski (1983), Cavanagh (1985), Chan and Wei (1987), Phillips (1987), and Stock (1991) that the asymptotic behavior of a t-statistic when the AR coefficient is local to unity is completely different from that for a fixed $|\rho| < 1$. If the autoregressive coefficient is local to unity , i.e. is defined by a sequence $\rho_T = \exp\{c/T\}$ for some fixed $c < 0$, then as the sample size increases we have convergence (7).

Phillips (1987) proved that

$$\left(\sqrt{-2c}\int_0^1 J_c(x)dw(x), (-2c)\int_0^1 J_c^2(x)dx\right) \Rightarrow (N(0,1),1) \quad \text{as} \quad c \to -\infty. \tag{14}$$

That is, we receive a classical normal approximation as a limiting case when c tends to negative infinity.

Let us consider a pair of statistics

$$(S^c, R^c) = \left(\frac{1}{\sqrt{g(c)}}\int_0^1 J_c^\mu(x)dw(x), \frac{1}{g(c)}\int_0^1 \left(J_c^\mu(x)\right)^2 dx\right),$$

where $J_c^\mu(x) = J_c(x) - \int_0^1 J_c(r)dr$, and $g(c) = E\int_0^1 (J_c^\mu(x))^2 dx$. We also define a function $c(T,\rho) = T\log(\rho)$. Stock's method suggests constructing an asymptotic confidence set as defined in (4) with $c_1(T,\rho)$ and $c_2(T,\rho)$ being $\alpha/2$ and $1-\alpha/2$

quantiles of the distribution of the statistic $\varphi_1 = \phi(S^{c(T,\rho)}, R^{c(T,\rho)}, T, \rho)$. An advantage of Stock's method is that the critical values depend on the one dimensional local parameter c and can be tabulated for commonly used levels of confidence and commonly used statistics.

Based on the construction the set $C(Y)$ has correct local to unity asymptotic coverage. Namely,

$$\lim_{T\to\infty} P_{\rho=\exp\{c/T\}}\left\{\rho \in C(Y)\right\} = 1-\alpha, \quad \forall c \leq 0.$$

The convergence (14) suggests that the method may work well for the values of the parameter ρ in the stationary region. However, until now there has been no proof of the uniform validity of Stock's method. We present this proof below.

Theorem 2 *Let Y be an AR(1) process with an intercept defined by model (1) with error terms satisfying the set of Assumptions A. Assume that the statistic $\varphi(Y,T,\rho) = \phi(S,R,T,\rho)$ belongs to the class $\mathcal{H}$. Let $C(Y)$ be a set defined by equation (4) with $c_1(T,\rho)$ and $c_2(T,\rho)$ being $\alpha/2$ and $1-\alpha/2$ quantiles of the distribution of the statistic $\varphi_1 = \phi(S^{c(T,\rho)}, R^{c(T,\rho)}, T, \rho)$. Then the set $C(Y)$ has an (uniform) asymptotic coverage probability $1-\alpha$.*

Remark 3 *The statement of Theorem 2 can be extended to an AR(1) process with a linear trend. Let*

$$(S^{\tau,c}, R^{\tau,c}) = \left(\frac{1}{\sqrt{g^\tau(c)}}\int_0^1 J_c^\tau(x)dw(x), \frac{1}{g^\tau(c)}\int_0^1 (J_c^\tau(x))^2 dx\right),$$

where $J_c^\tau(x) = J_c(x) - \int_0^1(4-6r)J_c(r)dr - x\int_0^1(12r-6)J_c(r)dr$, $g^\tau(c) = E\int_0^1(J_c^\tau(x))^2dx$. If Assumptions A are satisfied, the finite sample distribution of the statistic $\varphi = \phi(S^\tau, R^\tau, T, \rho)$ is uniformly approximated by the distribution of the statistic $\varphi_1 = \phi(S^{\tau,c(T,\rho)}, R^{\tau,c(T,\rho)}, T, \rho)$.

As we already proved in Theorem 1, the distribution of the statistic $\phi(S^N, R^N, T, \rho)$ provides a uniform approximation for the distribution of the statistic $\phi(S,R,T,\rho)$. In order to prove Theorem 2, it is enough to show that the distribution of the variable

$\phi(S^{c(T,\rho)}, R^{c(T,\rho)}, T, \rho)$ uniformly approximates the distribution of $\phi(S^N, R^N, T, \rho)$. It is easy to check all the conditions of Lemma 2 applied to pairs $(S^{c(T,\rho)}, R^{c(T,\rho)})$ and (S^N, R^N). The only non-trivial part is the validity of Condition 1, which is discussed below.

Let us consider a standard Brownian motion $w(t)$ and define the normal error terms by the following equality: $\frac{e_{T,j}}{\sqrt{T}} = w\left(\frac{j}{T}\right) - w\left(\frac{j-1}{T}\right)$. Then the AR(1) process $z_{T,t}(\rho)$ generated by $e_{T,j}$ has the following form:

$$\frac{z_{T,j}(\rho)}{\sqrt{T}} = \sum_{j=0}^{j} \rho^{j-i}\left(w\left(\frac{i}{T}\right) - w\left(\frac{i-1}{T}\right)\right) = \int_0^{\frac{j}{T}} e^{\log(\rho)(j-[Ts]-1)} dw(s).$$

Many statistics of interest can be represented as stochastic integrals. For example,

$$\frac{1}{\sqrt{g(T,\rho)}} \sum_{j=1}^{T} z_{j-1} e_j = \int_0^1 \int_0^t f_1(t, s, T, \rho) dw(s) dw(t),$$

where $f_1(t, s, T, \rho) = \frac{T}{\sqrt{g(T,\rho)}} e^{\log(\rho)([Tt]-[Ts]-1)} I\{s \leq \frac{[Tt]}{T}\}$. Its local to unity analogs has a similar form:

$$\frac{1}{\sqrt{g(c(T,\rho))}} \int_0^1 J_c(t) dw(t) = \int_0^1 \int_0^t f_2(t, s, T, \rho) dw(s) dw(t),$$

where $f_2(t, s, T, \rho) = \frac{1}{\sqrt{g(c(T,\rho))}} e^{\log(\rho)T(t-s)}$.

The lemma below says that the described statistics of interest are uniformly close to each other in the L_2 metric:

Lemma 5 *Let a set $\mathcal{A}_T$ be defined in (11). Then we have:*

a) $\lim_{T\to\infty} \sup_{\rho\in\mathcal{A}_T} E\left(\frac{1}{\sqrt{g(T,\rho)}} \sum_{j=1}^{T} z_{j-1} e_j - \frac{1}{\sqrt{g(c(T,\rho))}} \int_0^1 J_c(t) dw(t)\right)^2 = 0;$

b) $\lim_{T\to\infty} \sup_{\rho\in\mathcal{A}_T} E\left(\frac{1}{g(T,\rho)} \sum_{j=1}^{T} z_{j-1}^2 - \frac{1}{g(c(T,\rho))} \int_0^1 (J_c(t))^2 dt\right)^2 = 0;$

c) $\lim_{T\to\infty} \sup_{\rho\in\mathcal{A}_T} E\left(\frac{1}{\sqrt{g(T,\rho)}\sqrt{T}} \sum_{j=1}^{T} z_{j-1} - \frac{1}{\sqrt{g(c(T,\rho))}} \int_0^1 J_c(t) dt\right)^2 = 0;$

d) $\lim_{T\to\infty} \sup_{\rho\in\mathcal{A}_T} E\left(S^N - S^{c(T,\rho)}\right)^2 = 0;$

e) $\lim_{T\to\infty} \sup_{\rho\in\mathcal{A}_T} E\left(R^N - R^{c(T,\rho)}\right)^2 = 0.$

5 Validity of Hansen's method

The grid bootstrap was proposed by Hansen (1999) for AR(p) processes. This section considers a special case for the AR(1) model. The discussion of the general case will be presented in Section 7.

Let us consider a bootstrapped sample

$$y_t^* = \rho y_{t-1}^* + e_t^*, \quad y_0^* = 0, \quad e_t^* \sim \quad i.i.d.\ F_T,$$

where F_T is a distribution function, that can depend on Y. Let a pair of statistics $(S^*(T,\rho), R^*(T,\rho))$ be defined by

$$(S^*, R^*) = \left(\frac{1}{\sqrt{g(T,\rho)}} \sum_{t=1}^{T} y_{t-1}^{\mu,*} e_t^*, \frac{1}{g(T,\rho)} \sum_{t=1}^{T} (y_{t-1}^{\mu,*})^2 \right).$$

The grid bootstrap set $C(Y)$ is described by equation (4), where $\varphi = \phi(S, R, T, \rho)$, with $c_1(T,\rho)$ and $c_2(T,\rho)$ being $\alpha/2$ and $1-\alpha/2$ quantiles of the distribution of the statistic $\varphi_1 = \phi(S^*, R^*, T, \rho)$.

Hansen (1999) proposed that φ be a t-statistic and F_T be the cdf of the residuals. He proved that the distribution of φ_1 provides an asymptotic approximation of the distribution of φ for any fixed $|\rho| < 1$ and along a sequence of models with the local to unity AR coefficient $\rho_T = \exp\{c/T\}$. We prove that the grid bootstrap provides uniform approximation and constructs asymptotic confidence sets. We also generalize the procedure in two ways. First of all, we can use any test statistic from the class $\mathcal{H}$. Secondly, we allow for different specifications of the distribution function F_T. In particular, we can consider a parametric grid bootstrap, a non-parametric error based grid bootstrap and a non-parametric residual based grid bootstrap.

The sample Y is fully defined by the realized error terms $\Sigma_T = \{\varepsilon_j\}_{j=1}^T$ and the unknown true AR coefficient ρ. Assume that we are testing the hypothesis $H_0 : \rho = \rho_0$. Suppose that the distribution $F_T = F_T(Y, \rho_0)$ can depend on the sample Y and the null value ρ_0. Thus we can consider $F_T = F_T(\Sigma_T, \rho, \rho_0)$ as being a function of the realized error terms, the unknown true coefficient ρ and the null value ρ_0.

Definition Let $\mathcal{L}_r(K, M, \theta)$ be a class of sequences of distributions F_T satisfying the following 3 conditions:

1) $\mu_1(F_T) = 0$;

2) $\mu_2(F_T) = \sigma_T^2$, where $|\sigma_T^2 - 1| \leq MT^{-\theta}$;

3) $\sup_T |\mu|_r(F_T) < K$.

Here $\mu_j(F)$ is j-th central moment of distribution F, and $|\mu|_j(F)$ is j-th absolute moment of distribution F.

Theorem 3 *Let Y be an AR(1) process defined by equation (1) with error terms satisfying the set of Assumptions A. Assume that the statistic $\varphi(Y,T,\rho) = \phi(S,R,T,\rho)$ belongs to the class $\mathcal{H}$. Then the following three statements hold:*

1) $\lim_{T\to\infty}\sup_{\rho\in\Theta_T}\sup_{F_T\in\mathcal{L}_r(K,M,\theta)}\sup_x \left|P_\rho\{\varphi<x\} - P^*_\rho\{\varphi_1<x\}\right| = 0.$

2) If for almost all realizations of error terms $\Sigma = \{\varepsilon_1,...,\varepsilon_j,...\}$ there exist constants $K(\Sigma) > 0, M(\Sigma) > 0$ and $\theta > 0$ such that for all $\rho \in \Theta_T$ we have $F_T(\Sigma_T,\rho,\rho) \in \mathcal{L}_r(K,M,\theta)$, then

$$\lim_{T\to\infty}\sup_{\rho\in\Theta_T}\sup_x \left|P_\rho\{\varphi<x\} - P^*_\rho\{\varphi_1<x|\Sigma_T\}\right| = 0 \quad a.s.$$

That is, the bootstrap provides a uniform asymptotic approximation for almost all realizations of error terms.

3) Let the assumption from the second statement be satisfied. Let $C(Y)$ be a set defined by equation (4) with $c_1(T,\rho|Y)$ and $c_2(T,\rho|Y)$ being $\alpha/2$ and $1-\alpha/2$ quantiles of the distribution of the statistic $\varphi_1 = \phi(S^*,R^*)$ given the realization of Y. Then the set $C(Y)$ has asymptotic coverage probability $1-\alpha$.

In the rest of the section we discuss different choices of the bootstrap error distribution F_T. If F_T is taken from a parametric family, then the bootstrap is called parametric. We should note that Andrews' (1993) method is a version of the parametric grid bootstrap.

There are at least two ways of performing non-parametric grid bootstrap. The most intuitive one is to resample bootstrap errors from the residuals of the regression model (1). That is, let $\{\widehat{e}_j\}_{j=1}^T$ be residuals based on the sample Y: $\widehat{e}_j = y_j - \frac{1}{T}\sum_{i=1}^T y_i - \widehat{\rho} y^\mu_{j-1}$. The *residual based* bootstrap obtains error terms by resampling from $\{\widehat{e}_j\}_{j=1}^T$ with repetition. That is, $F_T^{res}(x) = \frac{1}{T}\sum_{j=1}^T I\{\widehat{e}_j \le x\}$. The distribution function F_T^{res} depends on the sample Y, but does not depend on the null hypothesis tested.

The second way of performing non-parametric grid bootstrap is to impose the null while finding the error terms. Suppose, that we are testing the null hypothesis

$H_0 : \rho = \rho_0$. Let us generate the sequence of error terms under the null $e_j(\rho_0) = y_j - \rho_0 y_{j-1}$. Note that if the null is true, then we have the true realization of unknown errors $e_j(\rho) = \varepsilon_j$. We recenter the errors $\widetilde{e}_j(\rho_0) = e_j(\rho_0) - \frac{1}{T}\sum_{i=1}^T e_i(\rho_0)$, and resample bootstrap errors from the centered errors: $F_T^{err}(x, \rho_0) = \frac{1}{T}\sum_{j=1}^T I\{\widetilde{e}_j(\rho_0) \le x\}$. We call this form of bootstrap *error based.* The distribution produced depends on the null value tested, and on the sample $F_T^{err}(x|Y, \rho_0) = F_T^{err}(x|\Sigma_T, \rho, \rho_0)$.

The lemma below states that the two non-parametric bootstrap procedures produce asymptotic confidence sets.

Lemma 6 *Assume that $\{y_t\}_{t=1}^T$ is a sample from an AR(1) process defined by (1) with errors satisfying the set of Assumptions A. Let $F_T^{res}(x|\Sigma_T, \rho)$ be an empirical distribution function for the residual based bootstrap and $F_T^{err}(x|\Sigma_T, \rho, \rho_0)$ be an empirical distribution function for the error based bootstrap. Then for every realization of errors Σ there exist constants $K(\Sigma) > 0, M(\Sigma) > 0$ and $\theta > 0$ such that for all $\rho \in \Theta_T$ we have $F_T^{res}(x|\Sigma, \rho) \in \mathcal{L}_r(K, M, \theta)$,and $F_T^{err}(x|\Sigma, \rho, \rho) \in \mathcal{L}_r(K, M, \theta)$.*

Remark 4 *Results of this section also hold for an AR (1) process with a linear trend, if all statistics are calculated for the detrended process in place of the demeaned.*

6 Why the Subsampling Procedure Fails

In order to construct a uniformly asymptotically valid confidence set it is sufficient to have a uniform asymptotic approximation. The subsampling procedure proposed by Romano and Wolf (2001) is aimed at constructing asymptotic confidence sets for the AR coefficient. They proved that the procedure is point-wise asymptotically correct. However, the sets provided by subsampling are not uniformly asymptotically correct. We are able to construct a sequence of AR(1) models with the AR coefficient depending on the sample size, such that the coverage probability of the set constructed by subsampling converges to a number lower than the declared coverage probability.

Let us consider a sample $\{z_j\}_{j=1}^T$ from the AR(1) process with intercept a and with i.i.d. normal innovations. If $|\rho| < 1$, the initial variable z_0 is normally distributed with mean $\frac{a}{1-\rho}$ and variance $\frac{1}{1-\rho^2}$. When $\rho = 1$, the initial value is an arbitrary constant. We base our inferences on the t-statistic $t(T,\rho)$ calculated from the sample of size T:

$$t(T,\rho) = \frac{\widehat{\rho}(T) - \rho}{\sigma(\widehat{\rho}(T))} = \frac{\sum_{j=1}^{T} e_j z_{j-1}^{\mu}}{\sqrt{\sum_{j=1}^{T}(z_{j-1}^{\mu})^2}}.$$

Let $b = b_T$ be a block size when the sample size is equal to T. We consider the subsample of size b starting from the observation j, that is, $\{z_j, z_{j+1}, ..., z_{j+b-1}\}$, and calculate the estimate $\widehat{\rho}_j(b)$, and the t-statistic using $\widehat{\rho}(T)$ as the null value:

$$\widehat{t}_j(b) = \frac{\widehat{\rho}_j(b) - \widehat{\rho}(T)}{\sigma(\widehat{\rho}_j(b))} = \frac{\sum_{i=j}^{j+b-1} e_i z_{i-1}^{\mu}}{\sqrt{\sum_{i=j}^{j+b-1}(z_{i-1}^{\mu})^2}} - (\widehat{\rho}(T) - \rho)\sqrt{\sum_{i=j}^{j+b-1}(z_{i-1}^{\mu})^2}.$$

Romano and Wolf (2001) argue that the unknown distribution of the t-statistic $t(T,\rho)$ could be well approximated by the empirical distribution function $L_{T,b}(x) = \frac{1}{T-b+1}\sum_{j=1}^{T-b+1} I\{\widehat{t}_j(b) \le x\}$. Let $q_{\alpha}^{L}(T,b)$ be the α- quantile of the distribution $L_{T,b}(x)$, then

$$C(T,b) = [\widehat{\rho}(T) - q_{1-\alpha/2}^{L}(T,b)\sigma(\widehat{\rho}(T)), \widehat{\rho}(T) - q_{\alpha/2}^{L}(T,b)\sigma(\widehat{\rho}(T))]$$

is the proposed equitailed confidence interval. Romano and Wolf (2001) proved that if $b_T \to \infty$ and $\frac{b_T}{T} \to 0$ as $T \to \infty$, then for every $\rho \in [0,1]$ the coverage probability of the interval $C(T,b)$ converges to $1-\alpha$, as the sample size increases.

Below we prove that subsampling is not a uniform procedure.

Theorem 4 *Let b_T be a sequence of natural numbers such that $b_T \to \infty$, and $\frac{b_T}{T} \to 0$ as $T \to \infty$. For any $c < 0$ set $\rho_T = 1 + c/b_T$, then*

$$\lim_{T\to\infty} P_{\rho_T}\{\rho_T \in C(T,b_T)\} < 1-\alpha.$$

Corollary 1 *The interval constructed by using subsampling is not an asymptotically uniform confidence set for the unrestricted parameter space $\Theta = (0,1)$.*

Romano and Wolf (2001) motivated their procedure by pointing out that the subsamples are generated from the same population as the whole sample, and as a result the autoregressive coefficients for the sample and the subsamples are the same.

This fact, according to them, should make quantiles of the distribution $L_{T,b}(x)$ close to the quantiles of the unknown distribution of the t-statistic $t(T, \rho)$.

However, the quality of approximations depends not only on the value of the autoregressive parameter ρ, but also on the sample size. Park (2003) notes that, the bigger sample you have, the wider is the range of ρ for which the normal approximation works well. The main idea of the proof of Theorem 4 lies in constructing a sequence of the coefficients ρ_T slowly converging to the unit root, such that the original sample size T is large enough and the limiting normal approximation is achieved, whereas the size of subsamples b_T is small and should be handled by the local to unity asymptotic approach.

One may be uncomfortable with this counterexample, which involves choosing ρ in response to the block size. In practice the block size is often data-driven. For example, we can choose the block size on the basis of the estimated persistence $\widehat{\rho}(T)$. This is the so called block size calibration method suggested by Romano and Wolf (2001, section 5.2). If the confidence interval is constructed by inverting hypothesis tests, we can choose different block sizes for different null hypotheses. In either case, the econometrician has the opportunity to choose b_T in response to the estimated or hypothesized value of ρ. However, even this flexibility in choosing the block size does not save the method.

Assume that we can choose the block size such that $b_{\min,T} \leq b_T(\rho) \leq b_{\max,T}$ with $b_{\min,T} \to \infty$ and $b_{\max,T}/T \to 0$. Let us consider $\rho_T = 1 + c/b_{\max,T}$ and $b_T = b_T(\rho_T)$. Since $(1 - \rho_T)T \to \infty$, the distribution of the test statistic calculated for the whole sample could be handled by the classical asymptotics. Let $(1 - \rho_{T_n})b_{T_n} \to \gamma$ be a converging subsequence of the sequence $\{(1 - \rho_T)b_T\}_{T=1}^{\infty}$. The distribution of the subsampled test statistics along this subsequence converges to the local to unity limiting distribution with the local to unity parameter γ, which is closer to 0 than c is. As a result, it is easy to show that

$$\limsup_{T\to\infty} P_{\rho_T}\{\rho_T \in C(T, b_T)\} < 1 - \alpha.$$

It is evident from the proof that similar results could be received for more general

AR(1) model with a linear trend. [4]

We should note that the result of Theorem 4 is true only for equi-tailed subsampling intervals. Symmetric subsampling intervals would have asymptotically uniformly correct coverage, but will be asymptotically uniformly conservative, i.e. the limit of the maximal coverage is higher than the declared level (see Andrews and Guggenberger (2005a)).

6.1 Small sample performance.

In this subsection we assess the extent to which the asymptotic results established in the paper are reflected in finite samples.

Romano and Wolf (2001) provided some Monte-Carlo simulations supporting subsampling. The main drawback of their results is that they considered a very restricted set of values of the AR coefficient, in particular, $\rho \in \{1, 0.99, 0.95, 0.9, 0.6\}$. They found that the subsampling works well for $\rho = 1$ and values of ρ very close to the unit root. Our asymptotic results predict that the subsampling intervals would have a good coverage for the unit root, but undercover for intermediate values of ρ, that is, for values which could be considered "stationary" for the whole sample, but "close to the unit root" for subsamples. Unfortunately, Romano and Wolf (2001) performed simulations for only one such value of the AR coefficient $\rho = 0.6$ (they used sample sizes $T = 120$ and $T = 240$). For $\rho = 0.6$ their 95% equitailed confidence intervals have a coverage probability of 77% , which is even worse than the coverage of the interval based on the normal approximation at this point.

Figure 1 shows a finite sample coverage of equitailed subsampling confidence intervals for a wide range of the AR coefficient in an AR(1) model with a linear time trend and normal errors. The sample size considered is $T = 120$. Subsampling intervals are constructed for block sizes $b = 5, 8, 12, 26$ (the grid suggested in Romano and Wolf

[4] Recently, Andrews and Guggenberger (2005b) independently showed that the subsampling may fail to provide asymptotically correct tests if used in a model where the limiting distribution of a test statistic is discontinuous in the true parameter. They consider inferences about autoregressive coefficient as one of examples.

(2001)).

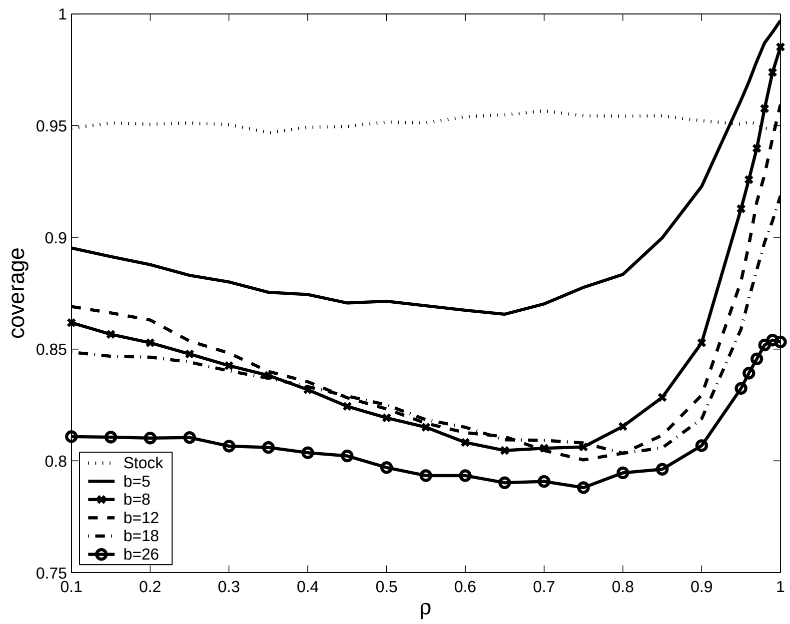

Figure 1. Coverage of equitailed interval constructed using local to unity asymptotics (Stock (1991)) and subsampling intervals with nominal level 95%. AR(1) model with a linear time trend, normal errors. Sample size = 120. Number of simulations =5000.

As expected, subsampling intervals undercover for all block sizes for quite a wide range of ρ. However, the extent of the problem is not as extreme as predicted by the asymptotic results of Andrews and Guggenberger (2005a). Results of additional simulations that can be found in the Supplementary Appendix show that the properties of subsampling intervals worsen as the sample size increases.

The method using local to unity asymptotics performs consistently well. In simulations we constructed three intervals advertised in this paper (Andrews's, Stock's and Hansen's). They all have coverage laying within simulation accuracy from the declared level over the whole parameter space. We depict only one of them, since all three lines are essentially indistinguishable. A more extensive simulation study can be found in the Supplementary Appendix on the author's web page.

7 AR(p) models.

This section extends the methods discussed in the previous sections to more empirically relevant AR(p) models. The proofs of all results from Section 7 have been placed in the Supplementary Appendix.

In this section we consider an AR(p) model with at most one root close to the unit circle. That is, we restrict all other roots to lay outside a circle strictly wider than the unit circle. Our aim is to make asymptotically uniformly correct inferences about the persistence of the series. There is a long discussion about the choice of a persistence measure in Andrews and Chen (1994). They provide arguments in favor of using the sum of the AR coefficients as opposed to the largest root. We concentrate our attention on the sum of the AR coefficients.

Let us consider an AR(p) model in ADF form:

$$y_t = \rho y_{t-1} + \sum_{j=1}^{p-1} \alpha_j \Delta y_{t-j} + \varepsilon_t, \tag{15}$$

where error terms satisfy Assumptions B.

Assumptions B. Let $\{\varepsilon_t\}_{t=1}^{\infty}$ be i.i.d. error terms with zero mean $E\varepsilon_t$=0, unit variance $E\varepsilon_t^2 = 1$ and a finite forth moment $E\varepsilon_t^4 < \infty$.

The process (15) can be described by equation $a(L)y_t = \varepsilon_t$, where $a(L) = 1 - \rho L - \sum_{j=1}^{p-1} \alpha_j (1-L)L^j$. Let us have the following representation of the polynomial $a(L) = (1-\mu_1 L)\cdot \dots \cdot (1-\mu_p L)$, where $|\mu_1| \leq |\mu_2| \leq \dots \leq |\mu_p| < 1$. Let us fix $0 < \delta < 1$. For every $\rho \in (0,1)$ we define a set $\mathcal{R}_\rho$ to be a set of all possible values of the nuisance parameter $\alpha = (\alpha_1, ..., \alpha_{p-1})$ for which $|\mu_{p-1}| < \delta$. It is easy to see the relationship between the sum of the AR coefficients, ρ, and the inverse roots $\{\mu_i\}_{i=1}^p$: $1 - \rho = (1-\mu_1)\cdot \dots \cdot (1-\mu_p)$. The case when ρ is close to one corresponds to μ_p being close to one. If $\rho = 1$, the process (15) has a unit root.

The main aim of this section is to construct an asymptotically uniformly correct confidence set for the parameter ρ. The procedure should work uniformly well for strictly stationary cases as well as in the situations when ρ is arbitrary close to 1. As before, the construction of a confidence set involves inverting a sequence of tests

$H_0 : \rho = \rho_0$.

We should note that a vector $\alpha = (\alpha_1, ..., \alpha_{p-1})$ is a nuisance parameter for the hypothesis $H_0 : \rho = \rho_0$. To test that the sum of the AR coefficients is equal to ρ_0 we calculate the conventional t-statistic $t(\rho_0, Y)$ for this hypothesis in the regression model (15). We also calculate $\widehat{\alpha}(\rho_0)$, an estimate of the nuisance parameter α, as the OLS estimator in the regression model with the null hypothesis imposed:

$$y_t - \rho_0 y_{t-1} = \sum_{j=1}^{p-1} \alpha_j \Delta y_{t-j} + \varepsilon_t. \tag{16}$$

That is, we regress $y_t - \rho_0 y_{t-1}$ on $\Delta y_{t-1}, ..., \Delta y_{t-p+1}$. Then we compare the calculated t-statistic $t(\rho_0, Y)$ with a critical value function $q(\rho_0, T, \widehat{\alpha}(\rho_0))$, depending on the tested value ρ_0 of the parameter of interest, on the estimated nuisance parameter, and on the sample size.

The confidence set for the parameter ρ is constructed as a set of values for which the corresponding hypothesis is accepted

$$C(y_1, ..., y_T) = \{\rho_0 : q_1(\rho_0, T, \widehat{\alpha}(\rho_0)) \leq t(\rho_0, Y) \leq q_2(\rho_0, T, \widehat{\alpha}(\rho_0))\}. \tag{17}$$

We consider two sets of critical value functions: the one received by parametric grid bootstrap, which is a generalization of Andrews' (1993) method, and those received by Hansen's (1999) non-parametric grid bootstrap. In the parametric grid bootstrap the critical value functions are quantiles of the distribution of the t-statistic $t(\rho_0, Z)$ in the model

$$z_t = \rho_0 z_{t-1} + \sum_{j=1}^{p-1} \widehat{\alpha}_j(\rho_0) \Delta z_{t-j} + e_t, \tag{18}$$

with errors e_t being independently normally distributed. In the non-parametric grid bootstrap we simulate critical value functions as quantiles of the distribution of the t-statistic in model (18) with i.i.d. error terms distributed according to a distribution function F_T. Below we prove the uniform asymptotic validity of both procedures.

7.1 Parametric grid bootstrap

When we have an AR(1) process with normal errors, the parametric grid bootstrap (Andrews' method) provides an exact confidence interval for the autoregressive coefficient ρ. The approximating distributions in AR(p) models employ the estimates of the nuisance parameter, rather than the true value of the nuisance parameter. As a result, the generalization of the method to AR(p) is not an exact method even if the error terms are normally distributed. We prove that the parametric grid bootstrap provides a uniform approximation of the unknown distribution of the t-statistic in an AR(p) model with normal errors as long as the estimate of the nuisance parameter is uniformly consistent.

Let statistics S and R be defined by

$$(S(Y,\rho,\alpha,T),R(Y,\rho,\alpha,T))=(G(\rho,\alpha)^{-1/2}\widetilde{Y}'\varepsilon,G(\rho,\alpha)^{-1/2}\widetilde{Y}'\widetilde{Y}G(\rho,\alpha)^{-1/2}),$$

where $\widetilde{Y}_t=(y_{t-1},\Delta y_{t-1},...,\Delta y_{t-p+1})$, $\widetilde{Y}=(\widetilde{Y}_1',...,\widetilde{Y}_T')'$, $\varepsilon=(\varepsilon_1,...\varepsilon_T)'$, and $G(\rho,\alpha)=diag\left(\sum_{t=1}^T Var(y_t),\sum_{t=1}^T Var(\Delta y_t),...,\sum_{t=1}^T Var(\Delta y_t)\right)$. Then the t-statistic for testing the hypothesis of the sum of AR coefficients being equal to ρ is

$$t(Y,\rho,\alpha,T)=l_1'R^{-1}(Y,\rho,\alpha,T)S(Y,\rho,\alpha,T)/\sqrt{l_1'R^{-1}(Y,\rho,\alpha,T)l_1},$$

where $l_1=(1,0,...,0)$.

Lemma 7 *Let us have two AR(p) processes: the process $Y=(y_1,...,y_T)$ defined by (15) and the process $Z=(z_1,...,z_T)$ defined by $z_t=\rho z_{t-1}+\sum_{j=1}^{p-1}\widehat{\alpha}_j\Delta z_{t-j}+\varepsilon_t$. Assume that error terms ε_j are the same for both processes and have i.i.d. standard normal distribution. Assume that the estimate $\widehat{\alpha}$ uniformly converges to α as the sample size increases*

$$\lim_{T\to\infty}\sup_{\rho\in[0,1)}\sup_{\alpha\in\mathcal{R}_\rho}P_\rho\left\{\|\alpha-\widehat{\alpha}\|>\epsilon\right\}=0 \quad \textit{for every} \quad \epsilon>0, \tag{19}$$

where $\|\alpha-\beta\|=\max_i|\alpha_i-\beta_i|$. Then

$$\lim_{T\to\infty}\sup_{\rho\in[0,1)}\sup_{\alpha\in\mathcal{R}_\rho}P_\rho\left\{|t(y,\rho,\alpha,T)-t(z,\rho,\widehat{\alpha},T)|>\varepsilon\right\}=0.$$

Hansen (1999) suggested to estimate the nuisance parameters by the OLS imposing the null. Lemma 8 states that the proposed estimates are uniformly consistent.

Lemma 8 *Assume that we have an AR(p) process defined by equation (15) with error terms satisfying the set of Assumptions B.*

Let us define $Y_t(\rho) = y_t - \rho y_{t-1}$, and $X_t = (\Delta y_{t-1}, .., \Delta y_{t-p+1})$. Let $\widehat{\alpha}$ be an OLS estimate in the regression of $Y_t(\rho)$ on X_t. Then $\widehat{\alpha}$ is a uniformly consistent estimate of α, that is, convergence (19) holds.

To prove that the parametric grid bootstrap is an asymptotically uniformly valid procedure for constructing confidence sets in models with non-normal errors, we employ the same idea as in Section 2. We divide the set of values of ρ into two overlapping subsets. One of the two subsets is increasing, while the second is contracting toward the unit root with a speed slower than $1/T$. The standard normal distribution provides a uniform approximation of the unknown distribution of the t-statistic over the first subset. We are able to construct two AR(p) processes with the same AR coefficients (one with normal errors, the other with errors ε_j) on a common probability space in such a way that the t-statistics for both processes are close to each other uniformly over the near unity set. As a result, the distribution of the t-statistic in an AR(p) model is uniformly approximated by the distribution of the t-statistic in an AR(p) model with the same AR coefficients but with normal errors. The validity of the parametric bootstrap procedure is stated in the theorem below.

Theorem 5 *Assume that $Y = (y_1, ..., y_T)$ is a sample from an AR(p) process defined by equation (15) with error terms satisfying the set of Assumptions B. Let $Z = (z_1, ..., z_T)$ be an AR(p) process with normal errors defined by equation (18), where $\widehat{\alpha}(\rho)$ is the OLS estimates in a regression model (16). Then the distribution of the t-statistic based on the sample Y can be uniformly approximated by the distribution of t-statistic based on the process Z:*

$$\lim_{T\to\infty} \sup_{\rho\in(0,1)} \sup_{\alpha\in\mathcal{R}_\rho} \sup_x |P\{t(Y,\rho,\alpha,T) > x\} - P\{t(Z,\rho,\widehat{\alpha}(\rho),T) > x\}| = 0.$$

As a result, the set defined by (17) with $q_i(\rho, T, \widehat{\alpha}(\rho)), i = 1, 2$ being quantiles of the distribution of $t(Z, \rho, \widehat{\alpha}(\rho), T)$, is a uniform asymptotic confidence set for ρ.

7.2 Non-parametric grid bootstrap

The non-parametric grid bootstrap procedure approximates the unknown distribution of the t-statistic $t(Y, \rho, \alpha, T)$ by the distribution of the t-statistic $t(Z, \rho, \widehat{\alpha}(\rho), T)$, where Z is an AR(p) process defined by (18) with error terms having distribution F_T. Let F_T be the empirical distribution function $F_T^{err}(\cdot)$ of the residuals from regression (16). Then $F_T(\Sigma, \rho_0, \rho, \alpha)$ depends on the realization of error terms of the process y_t, on the true coefficients ρ, α, and on the null value ρ_0 tested.

Theorem 6 *Assume that Y is a sample from an AR(p) process defined by equation (15) with error terms satisfying the set of Assumptions B. Let z_t be an AR(p) process defined by equation (18), where $\widehat{\alpha}(\rho)$ is the OLS estimates in a regression model (16). Assume that the errors e_t of the process z_t are i.i.d. with the distribution function F_T. Then the following three statements hold:*
1)

$$\lim_{T\to\infty} \sup_{\rho\in(0,1)} \sup_{\alpha\in\mathcal{R}_\rho} \sup_{F_T\in\mathcal{L}_r(K,M,\theta)} \sup_x |P\{t(Y,\rho,\alpha,T) > x\} - P\{t(Z,\rho,\widehat{\alpha},T) > x\}| = 0.$$

2) If for almost all realizations of error terms $\Sigma = \{\varepsilon_1, ..., \varepsilon_j, ...\}$ there exist constants $K(\Sigma) > 0, M(\Sigma) > 0$ and $\theta > 0$ such that $F_T(\Sigma, \rho, \rho, \alpha) \in \mathcal{L}_4(K, M, \theta)$, for all $\rho \in \Theta_T$, then

$$\lim_{T\to\infty} \sup_{\rho\in(0,1)} \sup_{\alpha\in\mathcal{R}_\rho} \sup_x \left|P_\rho\{t(Y,\rho,\alpha,T) > x\} - P_\rho^*\{t(Z,\rho,\widehat{\alpha},T) > x|\Sigma\}\right| = 0 \quad a.s.$$

That is, the bootstrap provides a uniform asymptotic approximation of the distribution of the t-statistic for almost all realizations of error terms.

c) Let $C(Y)$ be a set defined by equation (17) with $q_i(\rho, T, \widehat{\alpha}(\rho)) = q_i(\rho, T, \widehat{\alpha}(\rho)|Y), i = 1, 2$ being quantiles of the distribution of the statistic $t(Z, \rho, \widehat{\alpha}, T)$, where the bootstrapped errors e_t have the distribution function F_T^{err}. Then the set $C(Y)$ is an asymptotic confidence set.

8 Conclusion

In this paper I emphasize the difference between point-wise and uniform approximations. A point-wise approximation is not a strict enough condition for constructing an asymptotic confidence set, since it allows the convergence of the coverage probabilities to be extremely slow for some values of ρ. A uniform asymptotic approximation guarantees that we can achieve any accuracy uniformly over all possible values of ρ, as long as the sample size is large enough. Thus, having a uniform approximation of the unknown distribution of a test statistic always allows us to construct asymptotically valid confidence sets.

However, there still exists a common misleading belief in the literature that in order to construct a confidence set it is enough to check the validity of the procedure at each value of the parameter separately. Partially, it can be explained by the observation that the distinction between point-wise and uniform convergence is not important in many econometric applications (but it is here). We show the insufficiency of point-wise approximation by proving that Romano and Wolf's subsampling intervals are not asymptotic confidence sets, even though they are point-wise asymptotically correct.

This paper also fills a gap in the literature by proving the uniform validity of the three most used methods of constructing confidence sets for the persistence parameter in autoregressive models: Stock's local to unity method, Andrews' parametric grid bootstrap and Hansen's grid bootstrap.

Department of Economics, Harvard University, Littauer Center, Cambridge, MA 02138 USA; mikouch@fas.harvard.edu

APPENDIX

This Appendix provides proofs of the theorems and lemmas stated in Sections 2-6. Proof of the results from Section 7 are placed in the Supplementary Appendix, which can be found on the author's web-site.

Proof of Lemma 1.

$$P_\rho\{\rho \in C(Y)\} = F_{T,\rho}(q^G_{1-\alpha/2}(T,\rho)) - F_{T,\rho}(q^G_{\alpha/2}(T,\rho)) \geq$$

$$\geq G_{T,\rho}(q^G_{1-\alpha/2}(T,\rho)) - G_{T,\rho}(q^G_{\alpha/2}(T,\rho)) - 2\sup_x |F_{T,\rho}(x) - G_{T,\rho}(x)| =$$

$$= 1 - \alpha - 2\sup_x |F_{T,\rho}(x) - G_{T,\rho}(x)|$$

As a result,

$$\lim_{T\to\infty} \inf_{\rho\in\Theta} P_\rho\{\rho \in C(Y)\} \geq 1 - \alpha - 2\lim_{T\to\infty} \sup_{\rho\in\Theta} \sup_x |F_{T,\rho}(x) - G_{T,\rho}(x)| = 1 - \alpha.$$

Proof of Lemma 2.

$$\lim_{T\to\infty}\sup_{\rho\in\Theta_T}\sup_x|P_\rho\{\varphi<x\}-P_\rho\{\varphi_1<x\}|\le$$

$$\le\lim_{T\to\infty}\sup_{\rho\in\mathcal{A}_T}\sup_x|P_\rho\{\varphi<x\}-P_\rho\{\varphi_1<x\}|+\lim_{T\to\infty}\sup_{\rho\in\mathcal{B}_T}\sup_x|P_\rho\{\varphi<x\}-P_\rho\{\varphi_1<x\}|.$$

Let $\rho\in\mathcal{A}_T$. Then for the pairs (S,R) and (S_1,R_1) on a common probability space we have

$$P\{|\phi(S_1,R_1,T,\rho)-\phi(S,R,T,\rho)|>\varepsilon\}\le$$
$$\le P\{R_1<C\}+P\{R<C\}+P\{|S-S_1|+|R-R_1|>\tfrac{\varepsilon}{M_C}\}.$$

Here we use the fact that $\phi\in H$. Condition 4 of Lemma 2 allows one to choose $C>0$ such that $\sup_{\rho\in\Theta_T}P\{R_1<C\}$ is small enough. Since $P\{R<C\}<P\{R_1<2C\}+P\{|R-R_1|>C\}$, then due to Conditions 1 and 4 of the Lemma 2 we can make $\sup_{\rho\in\mathcal{A}_T}P\{R<C\}$ small enough. According to Condition 1, for a fixed C we can find T_1 such that $\sup_{\rho\in\mathcal{A}_T}P\{|S-S_1|+|R-R_1|>\frac{\varepsilon}{M_C}\}$ becomes small for $T>T_1$. As a result, the sequences of variables $\varphi_1=\phi(S_1,R_1,T,\rho)$ and $\varphi=\phi(S,R,T,\rho)$ converge to each other in probability uniformly over $\mathcal{A}_T$:

$$\lim_{T\to\infty}\sup_{\rho\in\mathcal{A}_T}P\{|\varphi_1(T,\rho)-\varphi(T,\rho)|>\varepsilon\}=0. \tag{20}$$

Note that,

$$P\{\varphi_1<x-\varepsilon\}-P\{|\varphi-\varphi_1|>\varepsilon\}\le P\{\varphi<x\}\le P\{\varphi_1<x+\varepsilon\}+P\{|\varphi-\varphi_1|>\varepsilon\}.$$

If the distribution of variable φ_1 is uniformly continuous

$$\lim_{T\to\infty}\sup\sup_{\rho\in\Theta_T}\sup_x P\{x-\varepsilon<\varphi_1(T,\rho)<x+\varepsilon\}\to 0 \ \text{ as } \ \varepsilon\to 0, \tag{21}$$

then statement (20) implies the closeness of distributions

$$\lim_{T\to\infty}\sup_{\rho\in\mathcal{A}_T}\sup_x|P_\rho\{\varphi<x\}-P_\rho\{\varphi_1<x\}|=0. \tag{22}$$

The statement that the distribution of variable φ_1 is uniformly continuous follows from the definition of the class $\mathcal{H}$ and Conditions 4 and 5 of Lemma 2. Namely, for every $0<C_1<C_2$ and $\delta>0$,

$$\begin{aligned}P\{x-\varepsilon<\phi(S_1,R_1,T,\rho)<x+\varepsilon\}&\le P\{R_1<C_1 \text{ or } R_1>C_2\}+\\+P\{x-\varepsilon<\phi(S_1,R_1,T,\rho)&<x+\varepsilon,|R_1-K|<\delta,C_1<R_1<C_2\}+\\+P\{x-\varepsilon<\phi(S_1,R_1,T,\rho)&<x+\varepsilon,|R_1-K|>\delta,C_1<R_1<C_2\}\end{aligned} \tag{23}$$

Condition 4 and Chebyshev's inequality imply $P\{R_1 > C_2\} \le \frac{K}{C_2}$. From this and Condition 4 it follows that the first summand of (23) can be made small enough by choosing large C_2 and small C_1.

According to the definition of H, $|\phi(S_1, R_1, T, \rho) - \phi(S_1, K, T, \rho)| \le M_{C_1}|R_1 - K|$:

$$P\{x - \varepsilon < \phi(S_1, R_1, T, \rho) < x + \varepsilon, |R_1 - K| < \delta, C_1 < R_1 < C_2\} \le$$
$$\le P\{x - \varepsilon - M_{C_1}\delta < \phi(S_1, K, T, \rho) < x + \varepsilon + M_{C_1}\delta\}.$$

Let us have $x - \varepsilon - M_{C_1}\delta < \phi(y, K, T, \rho) < x + \varepsilon + M_{C_1}\delta$ for some y. Since $\frac{\partial\phi(s,r,T,\rho)}{\partial s} > A > 0$ (see definition of H), we have $\phi(y + \Delta, K, T, \rho) > \phi(y, K, T, \rho) + A\Delta$ and $\phi(y - \Delta, K, T, \rho) < \phi(y, K, T, \rho) - A\Delta$ for all $\Delta > 0$. Thus,

$$P\{x - \varepsilon - M_{C_1}\delta < \phi(S_1, K, T, \rho) < x + \varepsilon + M_{C_1}\delta\} \le$$
$$\le P\left\{y - 2\frac{\varepsilon + M_{C_1}\delta}{A} < S_1 < y + 2\frac{\varepsilon + M_{C_1}\delta}{A}\right\} \le 2M\frac{\varepsilon + M_{C_1}\delta}{A}.$$

Here the last inequality follows from Condition 5 of Lemma 2. As a result, the second term in (23) can be made small by choosing small enough ε and δ.

Now we consider the last term in (23). Let $y = y(r_1)$ be a value such that $x - \varepsilon < \phi(y, r_1, T, \rho) < x + \varepsilon$ for some $C_1 < r_1 < C_2$. By using the same reasoning as above we receive the following ordering of invents $\{x - \varepsilon < \phi(S_1, r_1 + \delta, T, \rho) < x + \varepsilon\} \subseteq \{x - \varepsilon - M_{C_1}\delta < \phi(S_1, r_1, T, \rho) < x + \varepsilon + M_{C_1}\delta\} \subseteq \{y - 2\frac{\varepsilon + M_{C_1}\delta}{A} < S_1 < y + 2\frac{\varepsilon - M_{C_1}\delta}{A}\}$. By using continuity of the distribution of (S_1, R_1) (Condition 5 of Lemma 2) we get that

$$P\{x - \varepsilon < \phi(S_1, R_1, T, \rho) < x + \varepsilon, |R_1 - K| > \delta, C_1 < R_1 < C_2\} \le$$
$$\le \sum_k P\{x - \varepsilon < \phi(S_1, R_1, T, \rho) < x + \varepsilon, R_1 \in [x_k - \delta, x_k + \delta]\} \le$$
$$\le \sum_k P\{y(x_k) - 2\frac{\varepsilon + M_{C_1}\delta}{A} < S_1 < y(x_k) + 2\frac{\varepsilon - M_{C_1}\delta}{A}, R_1 \in [x_k - \delta, x_k + \delta]\} \le$$
$$\le \sum_k M2\frac{\varepsilon + M_{C_1}\delta}{A}\delta \le M\frac{\varepsilon + M_{C_1}\delta}{A}(C_2 - C_1) \le const(\varepsilon + M_{C_1}\delta),$$

here we divided a set $\{|R_1 - K| > \delta, C_1 < R_1 < C_2\}$ on intervals of the length 2δ. By choosing small ε and δ we can make the last term of (23) arbitrary small. It ends the proof of (21), and as a result, we have the uniform closeness of distributions (22).

Now let us consider $\rho \in \mathcal{B}_T$:

$$P\{\phi(S, R, T, \rho) < x\} \le P\{\phi(S, R, T, \rho) < x, |R - K| < \varepsilon\} + P\{|R - K| > \varepsilon\} \le$$
$$\le P\{\phi(S, K, T, \rho) < x + M_C\varepsilon\} + P\{|R - K| > \varepsilon\}$$

Similarly,

$$P\{\phi(S,R,T,\rho) < x - M_C\varepsilon\} \le P\{\phi(S,K,T,\rho) < x\} + P\{|R-K| > \varepsilon\}.$$

As a result,

$$P\left\{\phi(S,K,T,\rho) < x - M_C\varepsilon\right\} - P\{|R-K| > \varepsilon\} \le P\{\phi(S,R,T,\rho) < x\} \le \\ \le P\left\{\phi(S,K,T,\rho) < x + M_C\varepsilon\right\} + P\{|R-K| > \varepsilon\}. \tag{24}$$

According to Condition 3 of the Lemma, for any $\varepsilon > 0$ we can make $P\{|R-K| > \varepsilon\}$ arbitrary small uniformly over $\rho \in \mathcal{B}_T$. Function $s \mapsto \phi(s,K,T,\rho)$ is a continuous function uniformly with respect to (T,ρ) (definition of H). Condition 2 and the continuous mapping theorem imply that

$$\lim_{T\to\infty} \sup_{\rho\in\mathcal{B}_T} \sup_x |P\{\phi(S,K,T,\rho) < x\} - P\{\phi(\xi,K,T,\rho) < x\}| = 0, \tag{25}$$

where ξ has distribution $F(x)$. (24) and (25) imply that for any $\varepsilon > 0$ there is T_1 such that for all $T > T_1$

$$P\left\{\phi(\xi,K,T,\rho) < x - \varepsilon\right\} - \varepsilon \le P\{\phi(S,R,T,\rho) < x\} \le P\left\{\phi(\xi,K,T,\rho) < x + \varepsilon\right\} + \varepsilon. \tag{26}$$

Since $\phi(\cdot,K,T,\rho)$ is a continuous function and $F(x)$ is a continuous cdf,

$$P\{x - \varepsilon < \phi(\xi,K,T,\rho) < x + \varepsilon\} \to 0 \quad \text{as } \varepsilon \to 0.$$

As a result, (26) implies

$$\lim_{T\to\infty} \sup_{\rho\in\mathcal{B}_T} \sup_x |P\{\phi(S,R,T,\rho) < x\} - P\{\phi(\xi,K,T,\rho) < x\}| = 0.$$

The same statement is true for S_1. It gives us

$$\lim_{T\to\infty} \sup_{\rho\in\mathcal{B}_T} \sup_x |P\{\varphi < x\} - P\{\varphi_1 < x\}| = 0,$$

and completes the proof.

Lemma 9 *(A corollary of Theorem 2.18 from Hall and Heyde (1980)) Let ξ_j be a martingale difference sequence with $E|\xi_j|^\beta < \infty$ for some $1 < \beta < 2$. Then, for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} n^{-1/\beta-\varepsilon} \sum_{j=1}^{n} \xi_j = 0 \quad a.s.$$

Proof of Lemma 4. We start with a). It is easy to see that:

$$\begin{gathered}\tfrac{x_j}{\sqrt{T}} = \textstyle\sum_{i=1}^{j} \rho^{j-i} \left(\eta_T\left(\tfrac{i}{T}\right) - \eta_T\left(\tfrac{i-1}{T}\right)\right) = \\ = \textstyle\sum_{i=1}^{j}(\rho^{j-i} - \rho^{j-i-1})\eta_T\left(\tfrac{i}{T}\right) + \eta_T\left(\tfrac{j}{T}\right) - \rho^j \eta_T\left(\tfrac{0}{T}\right) = \\ = -(1-\rho)\textstyle\sum_{i=1}^{j} \rho^{j-i-1}\eta_T\left(\tfrac{i}{T}\right) + \eta_T\left(\tfrac{j}{T}\right)\end{gathered}$$

A similar expression is true for z_j: $\frac{z_j}{\sqrt{T}} = -(1-\rho)\sum_{i=1}^{j} \rho^{j-i-1} w_T\left(\frac{i}{T}\right) + w_T\left(\frac{j}{T}\right)$. So,

$$\sup_{\rho\in\Theta_T^+}\sup_j \left|\frac{x_j}{\sqrt{T}} - \frac{z_j}{\sqrt{T}}\right| \le$$

$$\le \sup_{\rho\in\Theta_T^+}\sup_j (1-\rho)\sum_{i=1}^{j}\rho^{j-i-1}\left|\eta_T\left(\frac{i}{T}\right) - w_T\left(\frac{i}{T}\right)\right| + \sup_j\left|\eta_T\left(\frac{j}{T}\right) - w_T\left(\frac{j}{T}\right)\right| \le$$

$$\le \sup_{\rho\in\Theta_T^+}\left(\frac{|1-\rho|}{\rho}\sup_j\sum_{i=1}^{j}\rho^{j-i} + 1\right)\sup_{0\le t\le 1}|\eta_T(t) - w_T(t)| = o(T^{-1/2+1/r+\varepsilon}).$$

b)

$$\sup_{\rho\in\Theta_T^+}\sup_{j=1,\dots,T}\left|\frac{x_j}{\sqrt{T}}\right| \le \sup_{\rho\in\Theta_T^+}\left(\frac{|1-\rho|}{\rho}\sup_{j=1,\dots,T}\sum_{i=1}^{j}\rho^{j-i} + 1\right)\sup_{0\le t\le 1}|w(t)| \le 2\sup_{0\le t\le 1}|w_T(t)|.$$

For c) we note that

$$\begin{gathered}\tfrac{1}{\sqrt{T}}\textstyle\sum_{j=1}^{T}\eta_T(j/T)\varepsilon_j = \textstyle\sum_{j=1}^{T}\left(\tfrac{\varepsilon_1}{\sqrt{T}} + \dots + \tfrac{\varepsilon_j}{\sqrt{T}}\right)\tfrac{\varepsilon_j}{\sqrt{T}} = \\ = \tfrac{1}{2}\left(\textstyle\sum_{j=1}^{T}\tfrac{\varepsilon_j}{\sqrt{T}}\right)^2 + \tfrac{1}{2T}\textstyle\sum_{j=1}^{T}(\varepsilon_j^2 - 1) + \tfrac{1}{2} = \tfrac{1}{2}(\eta_T(1))^2 + \tfrac{1}{2T}\textstyle\sum_{j=1}^{T}(\varepsilon_j^2 - 1) + \tfrac{1}{2}.\end{gathered}$$

According to Lemma 9, $\sum_{j=1}^{T}(\varepsilon_{T,j}^2 - 1) = o(T^{2/r+\varepsilon})\quad a.s.$ Consequently,

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{T}\eta_T(j/T)\varepsilon_j = \frac{1}{2}(\eta_T(1))^2 + \frac{1}{2} + o(T^{2/r-1+\varepsilon}).$$

By similar arguments,

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{T}w_T(j/T)e_{T,j} = \frac{1}{2}(w_T(1))^2 + \frac{1}{2} + o(T^{2/r-1+\varepsilon}).$$

As a result,

$$\left|\frac{1}{\sqrt{T}}\sum_{j=1}^{T}\eta_T(j/T)\varepsilon_j - \frac{1}{\sqrt{T}}\sum_{j=1}^{T}w_T(j/T)e_{T,j}\right| \le \frac{1}{2}\sup_{0\le t\le 1}(\eta_T(t) - w_T(t))^2 + o(T^{2/r-1+\varepsilon}) \le$$

$$\leq \sup_{0\leq t\leq 1}|\eta_T(t)-w(t)|(\sup_{0\leq t\leq 1}|\eta_T(t)|+\sup_{0\leq t\leq 1}|w(t)|)+o(T^{2/r-1+\varepsilon})=$$

$$=o(T^{-1/2+1/r+\varepsilon})+o(T^{2/r-1+\varepsilon})=o(T^{-1/2+1/r+\varepsilon}).$$

d) The formula for discrete integration by parts gives us the following:

$$\frac{1}{T}\sum_{j=1}^{T}x_{j-1}\varepsilon_j=\sum_{j=1}^{T}\left(\frac{x_{j-1}}{\sqrt{T}}-\frac{x_j}{\sqrt{T}}\right)\eta_T\left(\frac{j}{T}\right)+\frac{x_T}{\sqrt{T}}\eta_T\left(1\right).$$
$$\frac{1}{T}\sum_{j=1}^{T}z_{j-1}e_{T,j}=\sum_{j=1}^{T}\left(\frac{z_{j-1}}{\sqrt{T}}-\frac{z_j}{\sqrt{T}}\right)w_T\left(\frac{j}{T}\right)+\frac{z_T}{\sqrt{T}}w_T\left(1\right).$$

We note that $\frac{x_{j-1}}{\sqrt{T}}-\frac{x_j}{\sqrt{T}}=(1-\rho)\frac{x_{j-1}}{\sqrt{T}}-\frac{\varepsilon_j}{\sqrt{T}}$. As a result,

$$\frac{1}{T}\sum_{j=1}^{T}x_{j-1}\varepsilon_j-\frac{1}{T}\sum_{j=1}^{T}z_{j-1}e_{T,j}=(1-\rho)\sum_j\left(\frac{x_{j-1}}{\sqrt{T}}\eta_T\left(\frac{j}{T}\right)-\frac{z_{j-1}}{\sqrt{T}}w_T\left(\frac{j}{T}\right)\right)-$$
$$-\left(\frac{1}{\sqrt{T}}\sum_{j=1}^{T}\eta_T(j/T)\varepsilon_j-\frac{1}{\sqrt{T}}\sum_{j=1}^{T}w_T(j/T)e_{T,j}\right)+\left(\frac{x_T}{\sqrt{T}}\eta_T\left(1\right)-\frac{z_T}{\sqrt{T}}w_T\left(1\right)\right)$$

By applying a) and b) it is easy to see that

$$\sup_{\rho\in\Theta_T^+}\left|\frac{x_{j-1}}{\sqrt{T}}\eta_T\left(\frac{j}{T}\right)-\frac{z_{j-1}}{\sqrt{T}}w_T\left(\frac{j}{T}\right)\right|\leq\sup_{\rho\in\Theta_T^+}\sup_j\left|\frac{x_{j-1}}{\sqrt{T}}-\frac{z_{j-1}}{\sqrt{T}}\right|\sup_{0\leq t\leq 1}|w_T(t)|+$$

$$+\sup_{0\leq t\leq 1}|w_T(t)-\eta_T(t)|\sup_{\rho\in\Theta_T^+}\sup_j\left|\frac{x_{j-1}}{\sqrt{T}}\right|=o(T^{-1/2+1/r+\varepsilon}).$$

As a result, from c) we have

$$\sup_{\rho\in\Theta_T^+}\frac{1}{(1-\rho)T+1}\left|\frac{1}{T}\sum_{j=1}^{T}x_{j-1}\varepsilon_j-\frac{1}{T}\sum_{j=1}^{T}z_{j-1}e_{T,j}\right|=o(T^{-1/2+1/r+\varepsilon}).$$

Statements e) and f) can be obtained from a) and b):

$$\left|\frac{1}{T^2}\sum_{j=1}^{T}x_{j-1}^2-\frac{1}{T^2}\sum_{j=1}^{T}z_{j-1}^2\right|\leq$$

$$\leq\sup_j\left|\frac{x_j}{\sqrt{T}}-\frac{z_j}{\sqrt{T}}\right|\left(\sup_j\left|\frac{x_j}{\sqrt{T}}\right|+\sup_j\left|\frac{z_j}{\sqrt{T}}\right|\right)=o(T^{-1/2+1/r+\varepsilon});$$

$$\left|\frac{1}{T^{3/2+k}}\sum_{j=1}^{T}x_{j-1}j^k-\frac{1}{T^{3/2+k}}\sum_{j=1}^{T}z_{j-1}j^k\right|\leq$$

$$\leq\sup_j\left|\frac{x_j}{\sqrt{T}}-\frac{z_j}{\sqrt{T}}\right|\frac{1}{T^{k+1}}\sum_{j=1}^{T}j^k=o(T^{-1/2+1/r+\varepsilon}).$$

To check Statements g) and h) we use statements d), e) and f):

$$\sup_{\rho\in\Theta_T}\left|\frac{1}{T^2}\sum_{j=1}^{T}(y_{j-1}^\mu)^2-\frac{1}{T^2}\sum_{j=1}^{T}(z_{j-1}^\mu)^2\right|=\sup_{\rho\in\Theta_T}\left|\frac{1}{T^2}\sum_{j=1}^{T}x_{j-1}^2-\frac{1}{T^2}\sum_{j=1}^{T}z_{j-1}^2\right|+$$

$$+\sup_{\rho\in\Theta_T}\left|\left(\frac{1}{T^{3/2}}\sum_{j=1}^{T}x_{j-1}\right)^2-\left(\frac{1}{T^{3/2}}\sum_{j=1}^{T}z_{j-1}\right)^2\right|=o(T^{-1/2+1/r+\varepsilon})\quad a.s.;$$

$$\sup_{\rho\in\Theta_T}\frac{1}{(1+\rho)T+1}\left|\frac{1}{T}\sum_{j=1}^{T}y^{\mu}_{j-1}\varepsilon_j-\frac{1}{T}\sum_{j=1}^{T}z^{\mu}_{j-1}e_j\right|=$$

$$=\frac{1}{(1+\rho)T+1}\sup_{\rho\in\Theta_T}\left|\left(\frac{1}{T^{3/2}}\sum_{j=1}^{T}x_{j-1}\right)\eta_T(1)-\left(\frac{1}{T^{3/2}}\sum_{j=1}^{T}z_{j-1}\right)w_T(1)\right|+$$

$$+\frac{1}{(1+\rho)T+1}\sup_{\rho\in\Theta_T}\left|\frac{1}{T}\sum_{j=1}^{T}x_{j-1}\varepsilon_j-\frac{1}{T}\sum_{j=1}^{T}z_{j-1}e_j\right|=o(T^{-1/2+1/r+\varepsilon})\quad a.s.$$

We receive g) and h) from the two convergence statements above by noticing that $\sup_{\rho\in\mathcal{A}_T^+}\frac{T^2}{g(T,\rho)}=O(T^{1-\alpha})$. This completes the proof of Lemma 4.

Lemma 10 *(Corollary to Theorem 2 in Székely and Bakirov (2003)) For every $\varepsilon>0$ there exists $C>0$ such that*

$$\sup_T\sup_{\rho\in\Theta}P_\rho\left\{\frac{1}{g(T,\rho)}\sum_{t=1}^{T}(z^{\mu}_{t-1})^2<C\right\}<\varepsilon.$$

That is, the statistic R^N is uniformly separated from 0.

Proof of Theorem 1. One can check that the conditions of Lemma 2 are satisfied for sets $\mathcal{A}_T$ and $\mathcal{B}_T$ defined by (11) and (12) for $\frac{3}{4}+\frac{1}{2r}<\alpha<1$. Condition 1 follows from g) and h) of Lemma 4. Condition 4 is checked in Lemma 10.

Below we check Conditions 2 and 3. It is easy to see that

$$S(T,\rho)=\sqrt{\frac{T}{(1-\rho^2)g(T,\rho)}}\sqrt{\frac{1-\rho^2}{T}}\sum_{t=1}^{T}x_{t-1}\varepsilon_t-\left(\frac{1}{\sqrt{g(T,\rho)}}\sum_{t=1}^{T}x_{t-1}\right)\bar{\varepsilon},$$

$$R(T,\rho)=\frac{T}{(1-\rho^2)g(T,\rho)}\cdot\frac{1-\rho^2}{T}\sum_{t=1}^{T}x^2_{t-1}-\frac{1}{T}\left(\frac{1}{\sqrt{g(T,\rho)}}\sum_{t=1}^{T}x_{t-1}\right)^2.$$

Giraitis and Phillips (2006, Lemmas 2.1 and 2.2) proved that the following statements about convergence hold uniformly over $\mathcal{B}_T$:

$$\lim_{T\to\infty}\sup_{\rho\in\mathcal{B}_T}\sup_x\left|P\left\{\sqrt{\frac{1-\rho^2}{T}}\sum_{j=1}^{T}x_{j-1}\varepsilon_j\le x\right\}-\Phi(x)\right|=0,$$

$$\lim_{T\to\infty}\sup_{\rho\in\mathcal{B}_T} P\left\{\left|\frac{1-\rho^2}{T}\sum_{j=1}^{T} x_{j-1}^2 - 1\right| > \epsilon\right\} = 0.$$

Let us note that $\overline{\varepsilon} \to^p 0$. The fact that the term $\frac{1}{\sqrt{g(T,\rho)}}\sum_{t=1}^{T} x_{t-1}$ is bounded in probability uniformly over $\mathcal{B}_T$ can be shown by checking that its second moment is uniformly bounded. One also can check that

$$\lim_{T\to\infty}\sup_{\rho\in\mathcal{B}_T}\left|\frac{T}{(1-\rho^2)g(T,\rho)} - 1\right| = 0.$$

Combining all facts mentioned above, Conditions 2 and 3 of Lemma 2 are satisfied.

Proof of Theorem 2. According to Theorem 1 distribution of $\phi(S,R,T,\rho)$ is uniformly approximated by the distribution of $\phi(S^N,R^N,T,\rho)$. As a result, it is enough to check the conditions of Lemma 2 for two pairs of statistics $(S_1,R_1) = (S^N,R^N)$ and $(S,R) = (S^{c(T,\rho)}, R^{c(T,\rho)})$ and sets $\mathcal{A}_T$ and $\mathcal{B}_T$ defined by (11) and (12). Condition 1 follows from Lemma 5. Condition 4 has been checked in Lemma 10.

We check that Conditions 2 and 3 are satisfied. By simple calculations we receive that $\lim_{c\to-\infty}|-2cg(c)-1| = 0$, and $\lim_{c\to-\infty}(-2c)E\left(\int_0^1 J_c(r)dr\right)^2 = 0$. As a result, convergence (14) implies that as $c\to-\infty$

$$S^c = \sqrt{\frac{1}{-2cg(c)}}\left(\sqrt{-2c}\int_0^1 J_c(x)dw(x) - w(1)\sqrt{-2c}\int_0^1 J_c(r)dr\right) \Rightarrow N(0,1),$$

$$R^c = \frac{1}{-2cg(c)}(-2c)\int_0^1 J_c^2(x)dx - \frac{1}{-2cg(c)}\left(\sqrt{-2c}\int_0^1 J_c(r)dr\right)^2 \to^p 1.$$

Since $\lim_{T\to\infty}\max_{\rho\in\mathcal{B}_T} c(T,\rho) = -\infty$, Conditions 2 and 3 are satisfied for the pair $(S^{c(T,\rho)}, R^{c(T,\rho)})$.

Proof of Lemma 5. a) From the isomorphic property of Ito's integrals we have:

$$E\left(\frac{1}{\sqrt{g(T,\rho)}}\sum_{j=1}^{T} z_{j-1}e_j - \frac{1}{\sqrt{g(c(T,\rho))}}\int_0^1 J_c(t)dw(t)\right)^2 =$$
$$= \int_0^1\int_0^t (f_1(t,s,T,\rho) - f_2(t,s,T,\rho))^2 dsdt.$$

Let us introduce functions $f_3(t,s,T,\rho) = \frac{T}{\sqrt{g(T,\rho)}}e^{\log(\rho)([Tt]-[Ts]-1)}$ and $f_4(t,s,T,\rho) = \frac{T}{\sqrt{g(T,\rho)}}e^{\log(\rho)T(t-s)}$.

$$\int_0^1\int_0^t (f_1-f_2)^2 dsdt \le 2\int_0^1\int_0^t \left((f_1-f_3)^2 + (f_3-f_4)^2 + (f_4-f_2)^2\right) dsdt.$$

It is easy to see that

$$\lim_{T\to\infty}\sup_{\rho\in A_T}\int_0^1\int_0^t (f_1-f_3)^2 dsdt = \lim_{T\to\infty}\sup_{\rho\in A_T}\frac{T^2}{g(T,\rho)}\int_0^1\int_{\frac{[Tt]}{T}}^t e^{2\log(\rho)([Tt]-[Ts]-1)}dsdt =$$

$$= \lim_{T\to\infty}\sup_{\rho\in A_T}\frac{T}{g(T,\rho)}\rho^{-2} = 0.$$

For the second term we have:

$$\int_0^1\int_0^t (f_3-f_4)^2 dsdt = \frac{T^2}{g(T,\rho)}\int_0^1\int_0^t (e^{\log(\rho)([Tt]-[Ts]-1)} - e^{\log(\rho)T(t-s)})^2 dsdt \leq$$

$$\leq \frac{T^2}{g(T,\rho)}\int_0^1\int_0^t |\log(\rho)([Tt]-[Ts]-1)-\log(\rho)T(t-s)|^2 e^{2\log(\rho)([Tt]-[Ts]-1)}dsdt \leq 2\log^2(\rho).$$

We used the inequality $|e^{-a}-e^{-b}| \leq |a-b|e^{-a}$ that holds for $0<a<b$. As a result,

$$\lim_{T\to\infty}\sup_{\rho\in A_T}\int_0^1\int_0^t (f_3-f_4)^2 dsdt \leq \lim_{T\to\infty} T^{-2\alpha} = 0.$$

Finally, by simple calculation we can receive that

$$\lim_{T\to\infty}\sup_{\rho\in \mathcal{A}_T}\left|\frac{T^2 g(c(T,\rho))}{g(T,\rho)} - 1\right| = 0.$$

As a result, $\lim_{T\to\infty}\sup_{\rho\in A_T}\int_0^1\int_0^t (f_4-f_2)^2 dsdt = 0$. It completes the proof of a).

For b) we note that

$$E\left(\frac{1}{g(T,\rho)}\sum_{j=1}^T z_{j-1}^2 - \frac{1}{g(c(T,\rho))}\int_0^1 (J_c(t))^2\, dt\right)^2$$
$$= E\left(\int_0^1\left\{\left(\int_0^1 f_1(t,s)dw(s)\right)^2 - \left(\int_0^1 f_2(t,s)dw(s)\right)^2\right\}dt\right)^2 \leq$$
$$\leq E\int_0^1\left(\int_0^1 (f_1-f_2)\, dW_s\right)^2\left(\int_0^1 (f_1+f_2)\, dW_s\right)^2 dt \leq$$
$$= \sqrt{E\int_0^1\left(\int_0^1 (f_1-f_2)\, dW_s\right)^4 dt}\sqrt{E\int_0^1\left(\int_0^1 (f_1+f_2)\, dW_s\right)^4 dt}.$$

From Theorem 4 in Chapter 2 of Skorokhod (1965) we receive

$$E\left(\int_0^1 (f_1(t,s)-f_2(t,s))\, dW_s\right)^4 \leq 36\int_0^1 (f_1(t,s)-f_2(t,s))^4\, ds$$

and

$$E\left(\int_0^1 (f_1(t,s)+f_2(t,s))\, dW_s\right)^4 \leq 36\int_0^1 (f_1(t,s)+f_2(t,s))^4\, ds.$$

It is easy to check that

$$\sup_{\rho\in A_T}\int_0^1\int_0^1 (f_1(t,s)+f_2(t,s))^4\, dsdt = O(1) \quad \text{as} \quad T\to\infty.$$

The proof that $\lim_{T\to\infty}\sup_{\rho\in A_T}\int_0^1\int_0^1 (f_1(t,s)-f_2(t,s))^4\,dsdt = 0$ is completely analogous to that of part a). It finishes the proof of part b).

c)

$$E\left(\frac{1}{\sqrt{T}\sqrt{g(T,\rho)}}\sum_{j=1}^{T} z_{j-1} - \frac{1}{\sqrt{g(c(T,\rho))}}\int_0^1 J_{c(T,\rho)}(t)dt\right)^2 =$$

$$= E\left(\int_0^1\int_0^t f_1(s,t,T,\rho)dw(s)dt - \int_0^1\int_0^t f_2(s,t,T,\rho)dw(s)dt\right)^2 dt =$$

$$= \int_0^1\left(\int_s^1 f_1(s,t)dt - \int_s^1 f_2(s,t)dt\right)^2 ds \le \int_0^1\int_0^1 (f_1-f_2)^2\,dsdt \to 0$$

uniformly over $\mathcal{A}_T$ as $T\to\infty$.

d) and e). By simple algebraic manipulation we receive

$$S^N = \frac{1}{\sqrt{g(T,\rho)}}\sum_{j=1}^{T} z_{j-1}e_j - \left(\frac{1}{T^{1/2}}\sum_{j=1}^{T} e_{j-1}\right)\left(\frac{1}{\sqrt{g(T,\rho)}\sqrt{T}}\sum_{j=1}^{T} z_{j-1}\right);$$

$$S^c = \frac{1}{\sqrt{g(c)}}\int_0^1 J_c(t)dw(t) - w(1)\frac{1}{\sqrt{g(c)}}\int_0^1 J_c(t)dt,$$

and

$$R^N = \frac{1}{g(T,\rho)}\sum_{j=1}^{T} z_{j-1}^2 - \left(\frac{1}{\sqrt{g(T,\rho)}T^{1/2}}\sum_{j=1}^{T} z_{j-1}\right)^2;$$

$$R^c = \frac{1}{g(c)}\int_0^1 J_c^2(t)dt - \left(\frac{1}{\sqrt{g(c)}}\int_0^1 J_c(t)dt\right)^2.$$

Thus, statements a), b) and c) of Lemma 5 imply statements d) and e).

Lemma 11 *Let $\{\varepsilon_{T,j}; j=1,...,T; T\in\mathbb{N}\}$ be a triangular array of random variables, such that for every T variables $\{\varepsilon_{T,j}\}_{j=1}^{T}$ are i.i.d. with distribution F_T. Assume that $y_{T,j} = \rho y_{T,j-1} + \varepsilon_{T,j}$. Then for any sequence ρ_T such that $T(1-\rho_T)\to\infty$ we have*

$$\lim_{T\to\infty}\sup_{F_T\in\mathcal{L}_r(K,M,\theta)}\sup_{|\rho|\le\rho_T}\sup_x\left|P\left\{\frac{1}{\sqrt{g(T,\rho)}}\sum_{j=1}^{T} y_{T,j-1}\varepsilon_{T,j} < x\right\} - \Phi(x)\right| = 0,$$

and, for every $\epsilon > 0$,

$$\lim_{T\to\infty}\sup_{F_T\in\mathcal{L}_r(K,M,\theta)}\sup_{|\rho|\le\rho_T} P\left\{\left|\frac{1}{g(T,\rho)}\sum_{j=1}^{T} y_{T,j-1}^2 - 1\right| > \epsilon\right\} = 0;$$

$$\lim_{T\to\infty}\sup_{F_T\in\mathcal{L}_r(K,M,\theta)}\sup_{|\rho|\le\rho_T} P\left\{\left|\frac{1}{\sqrt{g(T,\rho)}\sqrt{T}}\sum_{j=1}^{T} y_{T,j-1}\right| > \epsilon\right\} = 0.$$

Lemma 12 *Let $\{\varepsilon_{T,j}; j = 1, ..., T; T \in \mathbb{N}\}$ be a triangular array of random variables, such that for every T the variables $\{\varepsilon_{T,j}\}_{j=1}^{T}$ are iid with cdf $F_T \in \mathcal{L}_r(K, M, \theta)$. Then we can construct a process $\eta_T(t) = \frac{1}{\sqrt{T}} \sum_{j=1}^{[Tt]} \varepsilon_{T,j}$ and a Brownian motions w_T on a common probability space in such a way that for every $\varepsilon > 0$ we have*

$$\lim_{T\to\infty} \sup_{F_T \in \mathcal{L}_r(K,M,\theta)} P\{\sup_{0\le t\le 1} |\eta_T(t) - w_T(t)| > \varepsilon T^{-\delta}\} = 0,$$

for some $\delta > 0$.

Proofs of Lemmas 11 and 12 can be found in the Supplementary Appendix.

Proof of Theorem 3. 1) Given the validity of Theorem 1 it is enough to prove that the bootstrapped statistics are uniformly approximated by statistics in a model with normal errors. We follow the framework proposed in Lemma 2. We check the conditions of Lemma 2 for two pairs of statistics $(S_1, R_1) = (S^N, R^N)$ and $(S, R) = (S^*, R^*)$ and sets $\mathcal{A}_T$ and $\mathcal{B}_T$ defined by (11) and (12). Condition 1 of Lemma 2 follows from Lemma 12 and the reasoning completely parallel to those in Lemma 4 with changing the speed of convergence from $-1/2 + 1/r$ to $-\delta$. Conditions 2 and 3 of Lemma 2 follow from Lemma 11. Conditions 4 and 5 were checked in the proof of Theorem 1.

Statement 2) of Theorem 3 trivially follows from statement 1).

3) From the definition of the grid bootstrap set we have that $\rho \in C(Y)$ if and only if $P^*_\rho\{\phi(S_1^*, R_1^*, T, \rho) > \phi(S, R, T, \rho)\ |F_T\} > \alpha/2$ and $P^*_\rho\{\phi(S_1^*, R_1^*, T, \rho) < \phi(S, R, T, \rho)\ |F_T\} > \alpha/2$. It is easy to see that

$$\sup_{\rho\in\Theta} P_\rho\{P^*_\rho\{\phi(S_1^*, R_1^*, T, \rho) > \phi(S, R, T, \rho)\ |F_T\} < \alpha/2\} \le$$

$$\le \sup_{\rho\in\Theta} P_\rho\{F(\phi(S, R, T, \rho), T, \rho) < \alpha/2 + \varepsilon\} +$$

$$+ \sup_{\rho\in\Theta} P_\rho \left\{ \sup_{\rho\in\Theta} \sup_{x} \left| P_\rho\{\phi(S, R, T, \rho) < x\} - P^*_\rho\{\phi(S_1^*, R_1^*, T, \rho) < x|\Sigma_T\} \right| > \varepsilon \right\},$$

where $F(x, T, \rho) = P_\rho\{\phi(S, R, T, \rho) < x\}$. The second term goes to zero for every $\varepsilon > 0$. The random variable $F(\phi(S, R, T, \rho), T, \rho)$ has a uniform distribution over $[0, 1]$, that is,

$$\sup_{\rho\in\Theta} P_\rho\{F(\phi(S, R, T, \rho), T, \rho) < \alpha/2 + \varepsilon\} = \alpha/2 + \varepsilon.$$

As a result, for every $\varepsilon > 0$ we have $\lim_{T\to\infty} \inf_{\rho\in\Theta} P_\rho\{\rho \in \Theta_T\} \geq 1 - (\alpha + 2\varepsilon)$. So, the coverage probability of $C(Y)$ converges to $1-\alpha$.

Proof of Lemma 6. First of all, we check that the residual based bootstrap produces F_T that belongs to $\mathcal{L}_r(K, M, \theta)$ class. The first condition of the class is trivially satisfied. For the third condition we have:

$$\frac{1}{T}\sum_{t=1}^{T} |\widehat{e}_t|^r \leq C_r \frac{1}{T}\sum_{t=1}^{T} |\varepsilon_t^\mu|^r + C_r \frac{1}{T}\sum_{t=1}^{T} |\widehat{e}_t - \varepsilon_t^\mu|^r,$$

where $\widehat{e}_j - \varepsilon_j^\mu = \frac{\sum_{i=1}^T \varepsilon_i y_{i-1}^\mu}{\sum_{i=1}^T (y_{i-1}^\mu)^2} y_{j-1}^\mu$.Let us consider each term separately. The first term is bounded a.s. due to the Strong Law of Large Numbers. For the second term we note that for every $\varepsilon > 0$ we have

$$\frac{1}{T}\sum_{j=1}^{T} \left|\widehat{e}_j - \varepsilon_j^\mu\right|^r = \frac{1}{T}\left|\frac{\sum_{j=1}^T \varepsilon_j y_{j-1}^\mu}{\sum_{j=1}^T (y_{j-1}^\mu)^2}\right|^r \sum_{j=1}^{T} |y_{j-1}^\mu|^r \leq \frac{1}{T}\left|\frac{\sum_{j=1}^T \varepsilon_j y_{j-1}^\mu}{\sum_{j=1}^T (y_{j-1}^\mu)^2}\right|^r \left(\sum_{t=1}^{T} |y_{t-1}^\mu|^2\right)^{r/2} =$$

$$= \frac{1}{T}\frac{\left|\frac{1}{\sqrt{g(T,\rho)}}\sum_{j=1}^T \varepsilon_j y_{j-1}^\mu\right|^r}{\left(\frac{1}{g(T,\rho)}\sum_{j=1}^T (y_{j-1}^\mu)^2\right)^{r/2}} = o_p(T^{-1+\varepsilon}).$$

Now, we check the second condition for the residual based bootstrap:

$$\frac{1}{T}\sum_{j=1}^{T} \widehat{e}_j^2 - 1 = \left(\frac{1}{T}\sum_{j=1}^{T} (\varepsilon_j^\mu)^2 - 1\right) + 3\frac{1}{T}\frac{\left(\frac{1}{\sqrt{g(T,\rho)}}\sum_{j=1}^T \varepsilon_j y_{j-1}^\mu\right)^2}{\frac{1}{g(T,\rho)}\sum_{j=1}^T (y_{j-1}^\mu)^2},$$

that converges a.s. to zero with a non-trivial speed since $E|\varepsilon_j|^r < \infty$ for $r > 2$.

For the error based bootstrap the errors under the null are $e_t(\rho) = \varepsilon_t$ the true errors. Then all conditions of $\mathcal{L}_r(K, M, \theta)$ class can be received directly from the Strong Law of Large Numbers.

Proof of Theorem 4. Let us fix a negative number c and define $\rho_T = 1 + \frac{c}{b_T}$. Then $T(1-\rho_T) = -\frac{c}{b_T}T \to \infty$. According to Park (2003) and Giraitis and Phillips (2006) we have the following convergence

$$\lim_{T\to\infty} \sup_x |P_{\rho_T}\{t(T, \rho_T) < x\} - \Phi(x)| = 0.$$

The t-statistics calculated for the sample that has only b_T observations follow the local to unity asymptotics with the local parameter c. In particular, from Lemma 13 of the Supplementary Appendix we have

$$\lim_{T\to\infty}\sup_x\left|P_{\rho_T}\{\widehat{t}_1(b)<x\}-P\left\{\frac{\int_0^1 K_c^\mu(x)dw(x)}{\sqrt{\int_0^1(K_c^\mu(x))^2dx}}<x\right\}\right|=0,$$

where $K_c^\mu(t)$ is the demeaned process $K_c(s)=J_c(s)+\frac{e^{cs}}{\sqrt{-2c}}\xi$, and $\xi\sim N(0,1)$ is independent of w. The difference between J_c and K_c is due to non-zero initial value of z_0, see more on that in Elliott (1999) and Elliott and Stock (2001).

According to Lemma 14 of the Supplementary Appendix we have

$$\lim_{T\to\infty}\sup_x\left|L_{T,b}(x)-P\left\{\frac{\int_0^1 K_c^\mu(s)dw(s)}{\sqrt{\int_0^1(K_c^\mu(s))^2ds}}<x\right\}\right|=0\ \text{ in probability.}$$

Since the cdf for the local to unity limit is a continuous function, the convergence above gives us the convergence of the quantiles

$$q_\alpha^L(T,b)-q_\alpha^c\to 0\ \text{ in probability.}$$

Here q_α^c denotes the $\alpha-$ quantile of the distribution of $\frac{\int_0^1 K_c^\mu(x)dw(x)}{\sqrt{\int_0^1(K_c^\mu(x))^2dx}}$. Finally, as shown in Section 6 of the Supplementary Appendix

$$\lim_{T\to\infty}P_{\rho_T}\{\rho_T\in C(T,b_T)\}=\Phi(q_{1-\alpha/2}^c)-\Phi(q_{\alpha/2}^c)<1-\alpha.$$

REFERENCES

Andrews D.W.K. (1993): " Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models," *Econometrica*, 61(1), 139-165.

Andrews D.W.K. and H.-Y. Chen (1994): " Approximately Median-Unbiased Estimation of Autoregressive Models," *Journal of Business and Economic Statistics*, 12(2), 187-202.

Andrews D.W.K. and P. Guggenberger (2005a): "Hybrid and Size-corrected Subsample Methods ," *unpublished manuscript*

Andrews D.W.K. and P. Guggenberger (2005b): "The Limit of Finite Sample Size and a Problem with Subsampling," *unpublished manuscript*

Bobkoski M.J. (1983): "Hypothesis testing in nonstationary time series," Unpublished PhD thesis (Dept. of Statistics, University of Wisconsin, Madison)

Cavanagh, C. (1985): “Roots local to unity”, Manuscript (Dept. of Economics, Harvard University, Cambridge, MA)

Chan N.H. and Wei C.Z. (1987): “Asymptotic Inference for Nearly Nonstationary AR(1) Processes,” *Ann. Statist.*, 15(3), 1050-1063.

Dickey, D.A. and W.A. Fuller (1979): “Distribution of the Autoregressive time series with a unit root,” *Journal of the American Statistical Society*, B, 50, 338-354.

Dufour J.-M. and M.L. King (1991): “Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary or Nonstationary AR(1) Errors,” *Journal of Econometrics*,47, 115-143.

Elliott G. (1999): “Efficient Tests for an Autoregressive Unit Root when the Initial Observation is Drawn from its Unconditional Distribution,” *International Economic Review* 40, 767-783.

Elliott G., T.J. Rothenberg and J.H. Stock (1996): “Efficient Tests for an Autoregressive Unit Root,” *Econometrica*, 64(4), 813-836.

Elliott G. and J.H. Stock (2001): “Confidence Intervals for Autoregressive Coefficients Near One,” *Journal of Econometrics*, 103, 155-181.

Hall P. and C.C. Heyde (1980): “Martingale Limit Theory and its Application,” *Academic Press, New York*

Hansen B.E. (1999): “The Grid Bootstrap and the Autoregressive Model,” *The Review of Economics and Statistics*, 81(4), 594-607.

Giraitis L, and Phillips P.C.B. (2006): “Uniform Limit Theory for Stationary Autoregression”, *Journal of Time Series Analysis* , 27(1), 51-60.

Imbs J., H.Mumtaz, M.O.Ravn, and H. Rey (2005): “PPP Strikes Back: Aggregation and the Real Exchange Rate,” *Quarterly Journal of Economics*, 120 (1), 1-43.

Lehmann E.L. (1997): “Testing Statistical Hypothesis,” *Springler Texts in Statistics*

Murray Ch.J. and D. H. Papell (2002): “The Purchasing Power Parity Persistence Paradigm,” *Journal of International Economics*, 56 (1), 1-19.

Nankervis J.C. and N.E. Savin (1985): “Testing the Autoregressive Parameter with the t-Statistic," *Journal of Econometrics*, 27, 143-161.

Nankervis J.C. and N.E. Savin (1988): "The Student's t Approximation in a Stationary First Order Autoregressive Model," *Econometrica*, 56(1), 119-145.

O'Reilly G. and K.Whelan (2005):"Has Euro-Area Inflation Persistence changed over time?" *The Review of Economics and Statistics*, 87(4), 709-720.

Park J.Y. (2003): "Weak Unit Roots," *unpublished manuscript.*

Phillips P.C.B. (1987): "Toward a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74(3), 535-547.

Rapach D.E. and M.E. Wohar (2004): "The Persistence in International Real Interest Rates," *International Journal of Finance and Economics*, 9 (4), 339-346.

Rayner R.K. (1990): "Bootstrapping p-Values and Power in the First-Order Autoregression: A Monte Carlo Investigation," *Journal of Business and Economic Statistic*, 8(2), 251-263.

Romano J.P. and Wolf M. (2001): "Subsampling Intervals in Autoregressive Models with Linear Time Trend," *Econometrica*, 69(5), 1283-1314.

Skorokhod A.V. (1965): "Studies in the Theory of Random Processes," *Addison-Wesley Publishing Company, Reading, Massachusetts*

Stock J. (1991): "Confidence Intervals for the Largest Autoregressive Root in US Macroeconomic Time Series," *Journal of Monetary Economics*, 28, 435-459.

Székely G.J. and Bakirov N.K. (2003): "Extremal Probabilities for Gaussian Quadratic Forms," *Probab. Theory Relat. Fields*, 126, 184-202.