



UvA-DARE (Digital Academic Repository)

Subsampling intervals in (un)stable autoregressive models with stationary covariates

van Giersbergen, N.P.A.

Publication date
2002

[Link to publication](#)

Citation for published version (APA):

van Giersbergen, N. P. A. (2002). *Subsampling intervals in (un)stable autoregressive models with stationary covariates*. (UvA Econometrics Discussion Paper; No. 2002/07). Department of Quantitative Economics.
[http://aimsrv1.fee.uva.nl/koen/web.nsf/view/38DDD63183ABA203C1256CF400237733/\\$file/0207.pdf](http://aimsrv1.fee.uva.nl/koen/web.nsf/view/38DDD63183ABA203C1256CF400237733/$file/0207.pdf)

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Discussion Paper: 2002/07

Subsampling intervals in (un)stable autoregressive models with stationary covariates

Noud P.A. van Giersbergen

www.fee.uva.nl/ke/UvA-Econometrics

Department of Quantitative Economics
Faculty of Economics and Econometrics
Universiteit van Amsterdam
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA  UNIVERSITEIT VAN AMSTERDAM



Subsampling Intervals in (Un)Stable Autoregressive Models with Stationary Covariates

Noud P.A. van Giersbergen
Department of Quantitative Economics,
University of Amsterdam
Roetersstraat 11
1018 WB Amsterdam
The Netherlands
Tel: +31-20-5254117
Fax: +31-20-5254349
E-mail: nvg@fee.uva.nl

October 9, 2002

Abstract

This paper considers confidence intervals based on the subsampling approach for the largest root in possibly unstable $AR(p)$ models with stationary exogenous regressors. The subsampling approach proposed by Politis and Romano (Annals of Statistics, 1994), is able to deal with discontinuities in the asymptotic distribution of a (studentized) estimator. We show that inference based on subsampling intervals is asymptotically correct. The finite-sample behavior of the subsampling approach is investigated by a small simulation study. It turns out that an equal-tailed subsampling interval based on a calibration rule leads to accurate inference in samples of size 100.

JEL classification: C12; C15; C22

Keywords: autoregressive models, confidence intervals, stationary covariates, subsampling, unit roots.

1 Introduction

The literature on the construction of asymptotically valid confidence intervals with accurate finite-sample performance in possibly unstable dynamic regression models is very limited due to the fact that the asymptotic distribution depends in a discontinuous way upon the largest root in the model. Although some progress has been made in the first-order autoregressive –AR(1)– model, see for instance Andrews (1993) and Stock (1991), to date, there exists no accurate inference procedure without specific distributional assumptions in general (un)stable dynamic regression models. The bootstrap, which was introduced two decades ago by Efron (1979), was hoped to be a general inference procedure with good small-sample properties. However, Basawa *et al.* (1991) have shown that confidence intervals based on the residual-based bootstrap fail in the unstable AR(1) model, i.e. the asymptotic coverage probabilities of the bootstrap confidence intervals do not converge to the desired nominal confidence level. In effect, the residual-based bootstrap fails because the AR(1) coefficient cannot be estimated with sufficient precision. By exploiting the duality between hypothesis testing and confidence intervals, Hansen (1999) has shown that valid confidence intervals for the AR(1) coefficient can be obtained by inverting a bootstrap test. However, his approach is quite specific and its validity depends on the fact whether the correct number of unit roots is imposed upon the model. It is only recently, that a general inference tool based on the idea of subsampling for weakly dependent (α -mixing) time series was introduced by Politis and Romano (1994). The main attraction of this subsampling approach is its ability to handle the earlier mentioned discontinuity of the asymptotic distribution. In addition, subsampling inference is consistent even in the presence of heteroskedasticity and (short-run) model misspecification; see e.g. Politis *et al.* (1997). Romano and Wolf (2001) extend the theory of subsampling to possibly unstable AR(p) models. From their analysis it appears that subsampling works if a properly normalized estimator has a non-degenerated asymptotic distribution and that the subsample statistics (thus not the data itself) are weakly dependent in some way. The purpose of this paper is to verify these requirements for the AR(p) model with stationary exogenous regressors allowing for a possible unit root in the estimation model. Such models are of considerable interest, because Hansen (1995) has shown that the power of unit root tests can be dramatically improved by including stationary covariates.

For ease of exposition, we shall focus on the AR(1) model including a univariate covariate in this section. Since the asymptotic distribution of the ordinary least-squares (OLS) estimator of the AR(1) coefficient depends on the deterministic variables included in the estimation model when there is a unit root, a constant and linear trend is always included in the model. Hence, we consider the following estimation model

$$y_t = \delta y_{t-1} + \beta_0 z_t + \mu + \pi t + \varepsilon_t, \quad t = 2, \dots, n, \quad (1)$$

where ε_t is i.i.d. with mean zero and variance σ_ε^2 , z_t is a stationary regressor and $\{\varepsilon_t\}$ is independent $\{z_t\}$. The goal is to construct a confidence interval for δ using only the observables $\{(y_t, z_t)\}_{t=1}^n$. Such a confidence interval can be based on an asymptotically pivotal quantity, which is a function of both the parameter of interest and its estimator. Here, the confidence interval is based on the studentized OLS estimator of δ , i.e.

$$\frac{\hat{\delta}_n - \delta}{SE(\hat{\delta}_n)}, \quad (2)$$

where $\hat{\delta}_n$ denotes the OLS estimator of δ and $SE(\hat{\delta}_n)$ denotes its estimated standard error. Although the model given in (1) seems very specific, a wealth of asymptotic distributions for (2) arises by varying the parameters (δ, β) . For instance, when $|\delta| < 1$, the standard normal distribution results, while a mixture of the Dickey-Fuller and standard normal is found when $\delta = 1$. Although outside the scope of this paper, even more distributions arise when z_t also contains a unit root. Then, both y_t and z_t are integrated of order 1 and the asymptotic distribution depends critically on the value of β_0 , i.e. if $\beta_0 = 0$ the variables do not cointegrate, while there is cointegration if $\beta_0 \neq 0$.

When $\delta = 1$ (and z_t is stationary), Hansen (1995) has shown that

$$\frac{\hat{\delta}_n - 1}{SE(\hat{\delta}_n)} \xrightarrow{d} \rho \frac{\int_0^1 \tilde{W} dW}{\left(\int_0^1 \tilde{W}^2 dr\right)^{1/2}} + (1 - \rho^2)^{1/2} \mathcal{N}(0, 1), \quad (3)$$

where W denotes a standard Brownian motion, \tilde{W} is a demeaned and detrended Brownian motion and ρ^2 denotes the long-run squared correlation between $\beta_0 z_t + \varepsilon_t$ and ε_t . For this particular model, this squared correlation is given by

$$\rho^2 = \frac{1}{1 + \left(\frac{\beta}{1-\gamma}\right)^2 \frac{\sigma_\beta^2}{\sigma_\varepsilon^2}}. \quad (4)$$

From (3), it follows that the asymptotic distribution is a convex combination of the Dickey-Fuller distribution and the standard normal. As $\rho^2 \rightarrow 1$, we find the Dickey-Fuller distribution, and as $\rho^2 \rightarrow 0$ we obtain the normal distribution. Hence, the asymptotic distribution depends on nuisance parameters, even asymptotically, which makes a resampling technique like the subsampling approach even more attractive. When $|\delta| < 1$, y_t and z_t are asymptotically stationary and by classic asymptotic theory we obtain

$$\frac{\hat{\delta}_n - \delta}{SE(\hat{\delta}_n)} \xrightarrow{d} \mathcal{N}(0, 1). \quad (5)$$

The aim of the paper is to analyze, both asymptotically and in small samples, the behavior of subsampling confidence intervals for the largest root in autoregressive models with stationary exogenous variables. Section 2 specifies the general model and its assumptions. In Section 3, the subsampling approach is briefly described, while Section 4 verifies the requirements for the subsampling method to work in our specific model. In Section 5, the subsampling confidence intervals are investigated by a small simulation study. Section 6 illustrates the calibration algorithm using empirical data. Finally, Section 7 concludes.

2 Model and Assumptions

Our framework largely corresponds to the framework considered in Hansen (1995). In unobserved-components form, the univariate series y_t is given by

$$y_t = \mu_y + \pi_y t + s_t. \quad (6)$$

The stochastic component s_t is modeled as

$$s_t = \delta s_{t-1} + \phi(L)\Delta s_t + v_t, \quad (7)$$

where $\phi(L) = \phi_1 L - \dots - \phi_p L^p$ is a p th order polynomial in the lag operator and

$$v_t = B(L)'(z_t - \mu_z) + \varepsilon_t. \quad (8)$$

In (8), z_t is an m -vector, $\mu_z = \mathbb{E}[z_t]$ and $B(L) = \beta_0 + \beta_1 L + \dots + \beta_q L^q$ is a lag polynomial of order q .

In regression form, the model for y_t is given by

$$y_t = \delta y_{t-1} + \phi(L)\Delta y_t + B(L)'z_t + \mu + \pi t + \varepsilon_t, \quad (9)$$

where $\mu = \mu_y(1 - \delta) + (\delta - \phi(1))\pi_y - B(1)'\mu_x$ and $\pi = \pi_y(1 - \delta)$. Note that the parameter π on the time trend is zero when $\delta = 1$ in (7) and (9). The asymptotic distribution of the OLS estimator $\hat{\delta}_n$ depends on the long-run squared correlation between v_t and ε_t . Hence, define the 2×2 long-run matrix

$$\Omega = \sum_{k=-\infty}^{\infty} \mathbb{E} \left[\begin{pmatrix} v_t \\ \varepsilon_t \end{pmatrix} \begin{pmatrix} v_{t-k} & \varepsilon_{t-k} \end{pmatrix} \right] = \begin{bmatrix} \sigma_v^2 & \sigma_{v\varepsilon} \\ \sigma_{v\varepsilon} & \sigma_\varepsilon^2 \end{bmatrix}$$

and

$$\rho^2 = \frac{\sigma_{v\varepsilon}^2}{\sigma_\varepsilon^2 \sigma_v^2}.$$

Of course, if there are no covariates, i.e. $B(L) = 0$, then clearly $\rho^2 = 1$. The opposite extreme, i.e. $\rho^2 = 0$, is excluded for technical reasons.

The covariates are assumed to be α -mixing. Specifically, given a random sequence $\{w_t, -\infty < t < \infty\}$, let \mathcal{F}_t^s be the σ -field generated by $\{w_t, \dots, w_s\}$, and define the corresponding α -mixing sequence by

$$\alpha_w(h) = \sup_t \sup_{A, B} |\mathbb{P}(A \text{ and } B) - \mathbb{P}(A)\mathbb{P}(B)|,$$

where A and B vary over the σ -fields $\mathcal{F}_{-\infty}^t$ and \mathcal{F}_{t+h}^∞ , respectively. The sequence $\{w_t\}$ is called α -mixing or strong mixing if $\alpha_w(h) \rightarrow 0$ as $h \rightarrow \infty$.

The assumptions are as follows:

Assumption 1 For some $r > 2$,

1. $\{z_t\}$ is covariance stationary and strong mixing with mixing coefficients $\alpha_z(h)$, which satisfy $\sum_{h=1}^{\infty} \alpha_z(h)^{1-2/r} < \infty$;
2. $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$
3. $\sup_t E[|z_t|^r + |\varepsilon_t|^r] < \infty$;
4. $E[z_{t-k}\varepsilon_t] = 0$ for all k ;

5. *the roots of $1 - \phi(z)$ all lie outside the unit circle;*
6. *$E[q_t q_t'] > 0$, where $q_t = (\Delta y_{t-1}, \dots, \Delta y_{t-p}, \Delta z_t' - \mu_z, \Delta z_{t-q}' - \mu_z)'$;*
7. *$\sigma_v^2 > 0$ and $\rho^2 > 0$.*

Discussion of these assumptions can be found in Hansen (1995). In contrast to Hansen, however, we assume that ε_t is uncorrelated with z_{t-k} for all k instead for $0 \leq k \leq q$. This is not a severe restriction since it is satisfied by any well-specified dynamic regression model. Moreover, it implies that $\rho^2 = \sigma_\varepsilon^2 / \sigma_v^2$. In addition, we assume that ε_t is i.i.d., while Hansen assumes that ε_t is α -mixing. Under assumption 1, the limiting distribution of the t -ratio of $\hat{\delta}_n$ is standard normal when $|\delta| < 1$ and a mixture of the Dickey-Fuller distribution and standard normal when $\delta = 1$.

3 The Subsampling Method

The purpose of this section is to summarize the general theory behind subsampling and it is heavily based on Romano and Wolf (2001, Section 3). Suppose $\{\dots, x_{-1}, x_0, x_1, \dots\}$ is a sequence of vector-valued random variables defined on a common probability space with joint probability distribution \mathcal{P} . The goal is to construct a confidence interval for some real-valued parameter $\theta = \theta(\mathcal{P})$ on the basis of a sample $\{x_1, \dots, x_n\}$. The time series $\{x_t\}$ will be assumed to satisfy a certain weak-dependence condition.

Let $\hat{\theta}_n = \hat{\theta}_n(x_1, \dots, x_n)$ be an estimator of $\theta(\mathcal{P}) \in \mathbb{R}^k$. Let $\hat{\theta}_{b,t} = \hat{\theta}_b(x_t, \dots, x_{t+b-1})$ denote the estimator of θ based on subsample $\{x_t, \dots, x_{t+b-1}\}$. Hence, the subindex b is the block size and t is the starting index of the block, so that $\hat{\theta}_{n,1} = \hat{\theta}_n$. We will focus on the studentized estimator $\tau_n(\hat{\theta} - \theta) / \hat{\sigma}_n$, where $\hat{\sigma}_n = \hat{\sigma}_n(x_1, \dots, x_n)$ is some positive estimator of scale. Define

$$\mathcal{L}_b(x) = \mathbb{P} \left(\tau_b(\hat{\theta}_{b,t} - \theta) / \hat{\sigma}_{b,t} \leq x \right) \quad (10)$$

as the distribution of $\tau_b(\hat{\theta}_{b,t} - \theta) / \hat{\sigma}_{b,t}$, which is assumed to be independent of t . Note that with this notation the finite-sample distribution of $\tau_n(\hat{\theta}_n - \theta) / \hat{\sigma}_n$ can be written as $\mathcal{L}_n(x)$. The major assumption that is needed to construct asymptotically valid confidence intervals for θ is the following.

Assumption 2 $\mathcal{L}_n(x)$ converges weakly to a non-degenerated distribution $\mathcal{L}(x)$. In addition, there exist positive sequences $\{a_n\}$ and $\{d_n\}$ such that $\tau_n = a_n/d_n$, $a_n(\hat{\theta}_n - \theta)$ converges weakly to a limit distribution $\mathcal{L}_1(x)$ and $d_n\hat{\sigma}_n$ converges weakly to a limit distribution $\mathcal{L}_2(x)$ without positive mass at zero.

Assumption 2 states that the estimator and its scale estimator, after being properly normalized, have a limiting distribution. It is hard to conceive of any asymptotic theory without such a requirement. The subsampling approximation to the exact sampling distribution $\mathcal{L}_n(x)$ is defined by

$$\hat{\mathcal{L}}_b(x) = \frac{1}{n-b+1} \sum_{t=1}^{n-b+1} 1\{\tau_b(\hat{\theta}_{b,t} - \hat{\theta}_n)/\hat{\sigma}_{b,t} \leq x\}. \quad (11)$$

The motivation behind the method is the following. The exact distribution of $\tau_b(\hat{\theta}_{b,t} - \theta)/\hat{\sigma}_{b,t}$ is $\mathcal{L}_b(x)$. If both b and n are large, then the empirical distribution of the $n-b+1$ values of $\tau_b(\hat{\theta}_{b,t} - \theta)/\hat{\sigma}_{b,t}$ should serve as a good approximation to $\mathcal{L}_n(x)$, which is the distribution we would like to approximate. Replacing the true parameter θ by its estimator $\hat{\theta}_n$ is permissible because $\tau_n(\hat{\theta}_n - \theta)$ is of order τ_b/τ_n in probability and we will assume that $\tau_b/\tau_n \rightarrow 0$. The following theorem, which is taken from Romano and Wolf (2001, Theorem 3.2), shows that confidence intervals for θ based on subsampling have asymptotically correct confidence coefficients.

Theorem 1 Under Assumption 2, suppose $a_b/a_n \rightarrow 0$, $\tau_b/\tau_n \rightarrow 0$, $b/n \rightarrow 0$ and $b \rightarrow \infty$ as $n \rightarrow \infty$. Also assume that $\{x_t\}$ is near epoch dependent of size $-q$, for some $q > 2$, on a basis process $\{\zeta_t\}$ whose α -mixing coefficients satisfy $\lim_{n \rightarrow \infty} n^{-1} \sum_{h=1}^n \alpha_\zeta(h)^{1-2/r} < \infty$ for some $r > 0$.

- (i) If x is a continuity point of $\mathcal{L}(x)$, then $\hat{\mathcal{L}}_b(x) \rightarrow \mathcal{L}(x)$ in probability.
- (ii) If $\mathcal{L}(x)$ is continuous, then $\sup_x |\hat{\mathcal{L}}_b(x) - \mathcal{L}(x)| \rightarrow 0$ in probability.
- (iii) For $\alpha \in (0, 1)$, let $\hat{q}_b(\alpha) = \inf\{x : \hat{\mathcal{L}}_b(x) \geq \alpha\}$. So, $\hat{q}_b(\alpha)$ serves as an α quantile of the subsampling distribution $\hat{\mathcal{L}}_b(x)$. Correspondingly, define $q(\alpha) = \inf\{x : \mathcal{L}(x) \geq \alpha\}$ as an α quantile of the asymptotic distribution $\mathcal{L}(x)$. If $\mathcal{L}(x)$ is continuous at $q(1-\alpha)$, then

$$\mathbb{P}\left(\tau_n(\hat{\theta} - \theta)/\hat{\sigma}_n \leq \hat{q}_b(\alpha)\right) \rightarrow 1 - \alpha \quad \text{as } n \rightarrow \infty.$$

Thus, the asymptotic coverage probability of the interval $I_1(b) = [\hat{\theta}_n - \hat{\sigma}_n/\tau_n \hat{q}_b(\alpha), \infty)$ is the nominal level $(1 - \alpha)$.

The interval $I_1(b)$ defined in Theorem 1 corresponds to a one-sided percentile- t interval in the bootstrap literature, see e.g. Hall (1992). A two-sided equal-tailed (ET) confidence interval is given by

$$I_2^{ET}(b) = [\hat{\theta}_n - \hat{\sigma}_n/\tau_n \hat{q}_b(1 - \alpha/2), \hat{\theta}_n - \hat{\sigma}_n/\tau_n \hat{q}_b(\alpha/2)]. \quad (12)$$

Although ET intervals could give important information about the asymmetry of a finite-sample distribution, Hall (1988) has shown that symmetric bootstrap confidence intervals enjoy enhanced coverage and can be shorter than equal-tailed confidence intervals, even in asymmetric circumstances. Hence, we also consider a two-sided symmetric (SYM) confidence interval for θ , which can be based on the sampling distribution

$$\mathcal{L}_n^{|\cdot|}(x) = \mathbb{P} \left(\tau_n |\hat{\theta}_n - \theta| / \hat{\sigma}_n \leq x \right). \quad (13)$$

Defining

$$\hat{\mathcal{L}}_b^{|\cdot|}(x) = \frac{1}{n - b + 1} \sum_{t=1}^{n-b+1} 1\{\tau_b |\hat{\theta}_{b,t} - \hat{\theta}_n| / \hat{\sigma}_{b,t} \leq x\} \quad (14)$$

as the subsampling approximation of $\mathcal{L}_n^{|\cdot|}(x)$, then a two-sided symmetric interval is given by

$$I_2^{SYM}(b) = [\hat{\theta}_n - \hat{\sigma}_n/\tau_n \hat{q}_b^{|\cdot|}(1 - \alpha), \hat{\theta}_n + \hat{\sigma}_n/\tau_n \hat{q}_b^{|\cdot|}(1 - \alpha)], \quad (15)$$

where $\hat{q}_b^{|\cdot|}(\alpha) = \inf\{x : \hat{\mathcal{L}}_b^{|\cdot|}(x) \geq \alpha\}$ denotes the α quantile of $\mathcal{L}_n^{|\cdot|}(x)$. Romano and Wolf (2001, Corollary 3.2) show that $I_2^{SYM}(b)$ has the correct asymptotic confidence level.

4 Subsampling in the AR(p) Model with Covariates

We shall now demonstrate that the general subsampling approach described in the previous section can be applied to construct asymptotically valid confidence intervals for the parameter δ in the estimation model

$$y_t = \delta y_{t-1} + \phi(L) \Delta y_t + B(L)' z_t + \mu + \pi t + \varepsilon_t. \quad (16)$$

Hence, δ will play the role of the general parameter θ of the previous section.

Let $\hat{\delta}_n$ denote the OLS estimator of δ . If \tilde{y}_t denotes the residual from regressing y_t on all the regressors shown in (16) except y_{t-1} , i.e.

$$\tilde{y}_t = y_t - \left(\sum_{j=t_0}^n y_j w'_j \right) \left(\sum_{j=t_0}^n w_j w'_j \right)^{-1} w_j, \quad (17)$$

where $t_0 = \max(p+1, q+1)$ and $w'_j = (\Delta y_{j-1}, \dots, \Delta y_{j-p}, z'_j, \dots, z'_{j-q}, 1, j)'$, then

$$\hat{\delta}_n - \delta = \frac{\sum \tilde{y}_{t-1} \varepsilon_t}{\sum \tilde{y}_{t-1}^2}. \quad (18)$$

Consequently, $\hat{\delta}_{b,t}$ is the OLS estimator for δ based on the block of data $\{y_t, \dots, y_{t+b-1}, z_t, \dots, z_{t+b-1}\}$. Denote the corresponding OLS standard errors by $SE(\hat{\delta}_n)$ and $SE(\hat{\delta}_{b,t})$. To apply the methodology of Section 2, we define $\hat{\sigma}_n = \sqrt{n} SE(\hat{\delta}_n)$ and $\hat{\sigma}_{b,t} = \sqrt{b} SE(\hat{\delta}_{b,t})$ and let $\tau_n = \sqrt{n}$ whatever the value of δ . The next theorem states that confidence intervals for δ based on the subsampling approach leads to consistent inference.

Theorem 2 *Assume that $b \rightarrow \infty$ and $b/n \rightarrow 0$ as $n \rightarrow \infty$ and that Assumption 1 holds. If $\theta = \delta$, $\hat{\theta}_n = \hat{\delta}_n$, $\hat{\sigma}_n = \sqrt{n} SE(\hat{\delta}_n)$, $x_t = (y_t z_t)$, and $\tau_n = \sqrt{n}$, then conclusions (i)–(iii) of Theorem 1 hold.*

Proof of Theorem 2. We have to show that the condition of Theorem 1 is satisfied for $\delta \in (-1, 1]$. Define $\zeta_t = (z'_t \varepsilon_t)'$. Consider first the case $|\delta| < 1$. Because $y_t = \mu_y + \pi_y t + s_t$, the residuals \tilde{y}_t are invariant with respect to (μ_y, π_y) . Hence, \tilde{y}_t only depends on $\{z_i, \varepsilon_i\}_{i=1}^t$, which is assumed to be covariance stationary and strong mixing. Furthermore, the estimation model is correctly specified so that ε_t is an innovation. By the Liapounov theorem for mixing processes, see for instance White (1984, Ch. 5), we obtain

$$\sqrt{n}(\hat{\delta}_n - \delta) = \frac{n^{-1/2} \sum \tilde{y}_{t-1} \varepsilon_t}{n^{-1} \sum \tilde{y}_{t-1}^2} \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma_\varepsilon^2}{\text{Var}(y_t)}\right) \quad (19)$$

and

$$\begin{aligned} \hat{\sigma}_n^2 &= n \hat{\text{Var}}(\hat{\delta}) \\ &= \frac{s_{\hat{\varepsilon}}^2}{n^{-1} \sum \tilde{y}_{t-1}^2} \xrightarrow{p} \frac{\sigma_\varepsilon^2}{\text{Var}(y_t)}. \end{aligned} \quad (20)$$

Thus Assumption 1 holds with $\tau_n = \sqrt{n}$, $a_n = \sqrt{n}$ and $d_n = 1$. Let $\alpha_{n,b}(\cdot)$ denote the α -mixing coefficients corresponding to the sequence $\{\tau_n(\hat{\delta}_{b,t} - \delta)/\hat{\sigma}_{b,t}, \hat{\sigma}_{b,t}\}_{t=1}^{n-b+1}$. It is sufficient to show that $n^{-1} \sum_{h=1}^n \alpha_{n,b}(h) \rightarrow 0$. To this end, note that both $\hat{\delta}_{b,t}$ and $\hat{\sigma}_{b,t}$ are functions of $(\tilde{y}_t, \dots, \tilde{y}_{t+b-1})$ so that $\alpha_{n,b}(h) \leq \min\{1, \alpha_\zeta(h-b)\}$. This implies the mixing condition $n^{-1} \sum_{h=1}^n \alpha_{n,b}(h) \rightarrow 0$ as $n \rightarrow \infty$, since $\alpha_\zeta(h) \rightarrow 0$ as $h \rightarrow \infty$ as well as $b/n \rightarrow 0$ as $n \rightarrow \infty$.

Next, consider the case $\delta = 1$. Let σ_v^2 denote the long-run variance of $v_t = B(L)'(z_t - \mu_z) + \varepsilon_t$. As shown in Hansen (1995),

$$n(\hat{\delta}_n - 1) \xrightarrow{d} (1 - \phi(1)) \frac{\sigma_\varepsilon}{\sigma_v} \left(\rho \frac{\int_0^1 \tilde{W} dW}{\int_0^1 \tilde{W}^2 dr} + (1 - \rho^2)^{1/2} \frac{\int_0^1 \tilde{W} dV}{\int_0^1 \tilde{W}^2 dr} \right) \quad (21)$$

and

$$\begin{aligned} \hat{\sigma}_n^2 &= n^2 \text{Vâr}(\hat{\delta}) \\ &= \frac{s_{\hat{\varepsilon}}^2}{n^{-2} \sum \tilde{y}_{t-1}^2} \xrightarrow{d} \frac{\sigma_\varepsilon^2}{\sigma_v^2 \int_0^1 \tilde{W}^2 dr}, \end{aligned} \quad (22)$$

where V is a standard Brownian motion independent of W .

To check the condition of Theorem 2, note that assumption 1 holds with $\tau_n = \sqrt{n}$, $a_n = n$ and $d_n = \sqrt{n}$. Next, we will show that $\hat{\delta}_{b,t} = \hat{\delta}_{b,t}(y_t, \dots, y_{t+b-1})$ and $\hat{\sigma}_{b,t} = \hat{\sigma}_{b,t}(y_t, \dots, y_{t+b-1})$ are functions of $(\zeta_t, \dots, \zeta_{t+b-1})$ only. First, note that we can always write $y_i = y_t + \sum_{s=t+1}^i \Delta y_s$ for $i \in \{t+1, \dots, t+b-1\}$. Hence, $\hat{\delta}_{b,t}$ and $\hat{\sigma}_{b,t}$ are functions of $(y_t, \Delta y_{t+1}, \dots, \Delta y_{t+b-1})$. When $\delta = 1$, we can always write ($i \in \{t+1, \dots, t+b-1\}$)

$$y_i = y_t + f(\Delta y_{t+1}, \dots, \Delta y_i, z_t, \dots, z_i, \varepsilon_{t+1}, \dots, \varepsilon_i), \quad (23)$$

so that y_i is a function of $(y_t, \Delta y_{t+1}, \dots, \Delta y_i, \zeta_t, \dots, \zeta_{t+b-1})$. However, the effect of the stochastic trend at time $t+1$, which is represented by y_t , is annihilated since the estimation model contains a constant. Hence, the residual \tilde{y}_i is not a function of y_t . Furthermore, the time series properties of Δy_i are fully determined by $\{z_s, \varepsilon_s\}_{s=t}^i$, so we only need to consider the time series properties $(\zeta_t, \dots, \zeta_{t+b-1})$. Since $\alpha_{n,b}(b+h) \leq \alpha_\zeta(h)$, we have that

$$\frac{1}{n} \sum_{h=1}^n \alpha_{n,b}(h) \leq \frac{b}{n} + \frac{1}{n} \sum_{h=1}^n \alpha_\zeta(h), \quad (24)$$

where the right-hand side converges to 0 by our mixing assumption on ζ_t . This completes the proof. \square

This section concludes with two remarks. First, the asymptotic validity of the subsampling approach only requires that the innovations ε_t are a martingale difference sequence. Hence, the assumption that ε_t are i.i.d. disturbances could be relaxed considerably. If one would maintain the hypothesis of martingale innovations, however, the assumption of a finite $r + 2$ moment of ε_t is required ($r > 2$). Secondly, as noted by Hansen (1995), the results stated in (21) and (22) depend critically on the assumption that z_t is stationary, i.e. $I(0)$. For instance, when both y_t and z_t are $I(1)$ but not cointegrated then the asymptotic distribution will be of the form found by Phillips and Ouliaris (1990). It is of considerable interest to see if the subsampling approach is robust with respect to the order of integration of the covariates.

5 Finite-sample Performance

In the previous section, the asymptotic validity of the subsampling approach was investigated. The asymptotic validity, however, is only a prerequisite. In this section, a small simulation study is performed to shed some light on the small-sample behavior of the subsampling method considered in this paper. All the simulations were carried out on a PC using the matrix programming language Gauss 3.2.

In the simulations, the data is generated according to an ARX(1) model given by

$$y_t = \delta y_{t-1} + \beta_0 z_t + \varepsilon_t, \quad t = 2, \dots, n, \quad (25a)$$

$$z_t = \gamma z_{t-1} + v_t, \quad (25b)$$

where $\varepsilon_t \sim \mathcal{NID}(0, 1)$, $v_t \sim \mathcal{NID}(0, 1)$ and $\{\varepsilon_t\}$ is independent of $\{v_t\}$. For $\delta \in \{0.6, 0.9, 0.95, 1\}$, $\gamma \in \{0.6, 0.9, 0.95\}$, $\beta \in \{0.0, 0.1, 0.4\}$, $N = 1000$ realizations of $\{y_t, z_t\}$ were generated according to the DGP given in (25) with sample size $n = 100$. We consider equal-tailed (ET) as well as symmetric (SYM) confidence intervals with a nominal confidence coefficient of 95%.

The performance of subsampling intervals critically depends on the ability to choose the ‘‘right’’ block size. This is due to the fact that for b close to n , all subsample statistics $\hat{\delta}_{b,t}$ will almost equal the estimate $\hat{\delta}_n$ obtained from the whole sample. If b is too small, however, the intervals can undercover or overcover depending on the time-series properties. The correct block size, say b^* , is chosen according to one of the two algorithms proposed by Romano and

Wolf (2001, Algorithms 5.1 and 5.2). The first algorithm is based on minimizing a running standard deviation for the confidence limits. The formal description is as follows:

1. For each b contained in the interval $[b_{small}, b_{big}]$ compute a subsampling interval for δ denoted by $I(b)$.
2. Compute for each b a volatility index $VI(b)$ as the standard deviation of the confidence limits over the interval $[b-k, b+k]$. The constant k in the algorithm denotes a “window” parameter. For both types of intervals considered (equal-tailed and symmetric), the standard deviation is calculated with respect to the lower confidence limits $\{I_{low}(b-k), \dots, I_{low}(b+k)\}$ and the upper confidence limits $\{I_{high}(b-k), \dots, I_{high}(b+k)\}$.
3. Choose the optimal subsample size b^* that minimizes $VI(b)$ and report the interval based on the subsample size b^* and the original data, i.e. $I(b^*)$.

Secondly, we have investigated the performance of the calibration algorithm for choosing the “optimal” block size b^* . This algorithm can be described as follows:

1. Generate K residual-based bootstrap replications of y_t based on

$$y_t = \hat{\delta}y_{t-1} + \hat{\beta}_0 z_t + \hat{\mu} + \hat{\beta}t + \hat{\varepsilon}_t. \quad (26)$$

Let $\mathcal{Y}_k^* = \{y_{1,k}^*, \dots, y_{T,k}^*\}$ denote the k -th bootstrap sample.

2. For each bootstrap sample \mathcal{Y}_k^* and block size $b \in [b_{small}, b_{big}]$, calculate the subsample confidence interval for δ denoted by $I_k^*(b)$.
3. Compute the bootstrap estimate of the confidence level for each b , i.e.

$$1 - \hat{\alpha}^*(b) = \frac{1}{K} \sum 1\{\hat{\delta}_n \in I_k^*(b)\},$$

where $1\{\cdot\}$ denotes the indicator function and $\hat{\delta}_n$ is the estimate of δ based on the original sample.

4. Choose the optimal subsample size b^* such that the difference between the estimated confidence level based on the bootstrap replications and the desired nominal level are minimal, i.e.

$$b^* = \arg \min_b (\alpha - \hat{\alpha}(b))^2.$$

5. The final interval is the subsampling interval based on the subsample size b^* and the original data, i.e. $I(b^*)$.

It is noted that for the ET confidence interval, separate optimal block sizes are determined for the lower confidence limit (LCL) and the upper confidence limit (UCL) in both algorithms.

In the Monte Carlo study, we set $k = 2$ and choose $(b_{small}, b_{big}) = (8, 27)$. Tables 1 (based on the first algorithm) and 2 (based on the second algorithm) show the estimates of the coverage probabilities in the estimation model including a constant and linear trend; see equation (26).

Insert Tables 1 and 2 about here.

From Table 1, which shows the results when minimizing the volatility, we observe that the ET interval undercovers the true parameter significantly. Furthermore, the coverage probabilities have a tendency to rise as the AR(1) parameter δ increases towards 1. The effect of γ on the coverage probabilities seems only marginal. The SYM intervals significantly outperform the ET intervals except when $\delta = 1$. This can be explained by the fact that when $\delta = 0$ and β is small, the finite-sample distribution of $\hat{\delta}_n$ is asymmetric. Unreported simulation results show that the coverage probabilities decrease as b increases. Hence, we conclude that the minimizing volatility algorithm chooses a block size which is larger than the “right” block size.

From Table 2, which gives the result for the calibration algorithm, we observe that subsampling inference can be very accurate. On average, equal-tailed intervals is at least as good as symmetric intervals. However, when $\delta = 1$, the ET intervals significantly outperforms the symmetric intervals.

Overall we conclude that the finite-sample performance of the subsampling is highly depend on the algorithm which is used to chose the “right” block size. The simulation results show that the calibration algorithm is superior to the volatility algorithm. This latter algorithm does not lead to accurate subsampling inference, at least in the parametrizations we have considered. The calibration algorithm, however, leads to estimated coverage probabilities which are very close to the nominal confidence level. The feasibility of this algorithm in practice is shown in the next section.

6 Empirical Illustration

In this section, we shall construct a equal-tailed 95% subsampling interval based on the calibration algorithm for the largest root in industrial production using the first differences of unemployment as covariate. We consider annual data for the U.S. from 1890 through 1988 as used by Nelson and Plosser (1982) and extended by Schotman and van Dijk (1991). The log of industrial production was chose as the dependent variable (y_t), while the first differences of unemployment (no logs) was used as a covariate (z_t). The estimation equation includes a constant, linear time trend and three lags of the first differences. OLS yields the following estimates (t -values between parentheses):

$$y_t = \begin{array}{cccc} 0.941 y_{t-1} & - 0.165 \Delta y_{t-1} & - 0.050 \Delta y_{t-2} & - 0.037 \Delta y_{t-3} \\ (-1.43) & (-2.61) & (-0.87) & (-0.64) \\ - 0.035 z_t & + 0.128 & + 0.002t & + \hat{\varepsilon}_t, \quad t \in \{1895, \dots, 1988\} \end{array} \quad (27)$$

In the calibration algorithm, we have to estimate the coverage probability for the lower confidence limit based on the 97.5% quantile and for the upper confidence limit based on the 2.5% quantile. These estimates are calculated from 1,000 bootstrap samples based on equation (27) and they are shown in Figures 1 and 2. From these figures, it can be read off that the optimal block sizes for the 2.5% and 97.5% quantiles are respectively 21 and 11. Hence, the optimal block size for the upper confidence limit is almost two times as large as the optimal block size for the lower confidence limit. This choice finally leads to the following 95% subsampling interval for the largest root in the model

$$(0.941 - 0.042 \cdot 2.016, 0.941 + 0.042 \cdot 3.454) = (0.856, 1.086). \quad (28)$$

This interval clearly contains the value 1, so that the hypothesis of a unit root cannot be rejected.

Insert Figures 1 and 2 about here.

7 Conclusion

Based on the theory developed in Romano and Wolf (2001), we have shown that subsampling inference is valid for the largest root in possibly unstable $AR(p)$ models with stationary ex-

ogenous regressors, i.e. subsampling confidence intervals for δ have the correct confidence level in the model

$$y_t = \delta y_{t-1} + \phi(L)\Delta y_t + B(L)'z_t + \mu + \pi t + \varepsilon_t, \quad (29)$$

whether $|\delta| < 1$ or $\delta = 1$. Although classical confidence intervals for δ could be obtained by pretesting procedures, such procedures are hampered by small-sample problems. By re-computing the studentized OLS estimator of δ , the subsampling approach leads to a natural finite-sample approximation which is able to handle discontinuities of the limiting distribution of the estimator as a function of the underlying model parameters.

In our simulation study, two rules for selecting the correct block size are considered. The minimum volatility rule leads to estimated coverage probabilities which are uniformly too low. Subsampling inference based on the calibration rule turns out to be accurate in samples with 100 observations, at least in the parametrizations we have considered. The equal-tailed interval outperforms the symmetric interval in case a unit root is present.

To conclude, the subsampling approach is a general and powerful technique, which can yield accurate inference in model (29) under very general assumptions. The interesting question whether subsampling inference is robust with respect to the order of integration of the covariates is currently under investigation.

References

- Andrews, D. W. K. (1993). Exactly Median-unbiased Estimation of First-order Autoregressive/ Unit-root Models. *Econometrica*, **61**, 139–165.
- Basawa, I. V., A. K. Mallik, W. P. McCormick, J. H. Reeves, and R. L. Taylor (1991). Bootstrapping Unstable First Order Autoregressive Processes. *The Annals of Statistics*, **19**, 1089–1101.
- Efron, B. (1979). Bootstrap Methods: Another look at the Jackknife. *The Annals of Statistics*, **7**, 1–26.
- Hall, P. (1988). On Symmetric Bootstrap Confidence Intervals. *Journal of the Royal Statistical Society, series B*, **50**, 35–45.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer-Verlag.
- Hansen, B. E. (1995). Rethinking the Univariate Approach to Unit Root Testing. *Econometric Theory*, **11**, 1148–1171.
- Hansen, B. E. (1999). The Grid Bootstrap and the Autoregressive Model. *The Review of Economics and Statistics*, **81**, 594–607.
- Nelson, C. R. and C. I. Plosser (1982). Trends and Random Walks in Macroeconomic Time Series. *Journal of Monetary Economics*, **10**, 139–162.
- Phillips, P. C. B. and S. Ouliaris (1990). Asymptotic properties of Residual Based Tests for Cointegration. *Econometrica*, **58**, 165–193.
- Politis, D. N. and J. P. Romano (1994). Large Sample Confidence Regions Based on Subsampling under Minimal Assumptions. *The Annals of Statistics*, **22**, 2031–2050.
- Politis, D. N., J. P. Romano, and M. Wolf (1997). Subsampling for Heteroskedastic Time Series. *Journal of Econometrics*, **81**, 281–317.
- Romano, J. P. and M. Wolf (2001). Subsampling Intervals in Autoregressive Models with Linear Time Trend. *Econometrica*, **69**, 1283–1314.
- Schotman, P. C. and H. K. van Dijk (1991). On Bayesian Routes to Unit Roots. *Journal of Applied Econometrics*, **6**, 387–401.
- Stock, J. H. (1991). Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series. *Journal of Monetary Economics*, **28**, 435–459.
- White, H. (1984). *Asymptotic Theory for Econometricians*. Orlando: Academic Press.

Figure 1: Estimated coverage probability as function of the blocksize b based on 1,000 residual-based bootstrap replications.

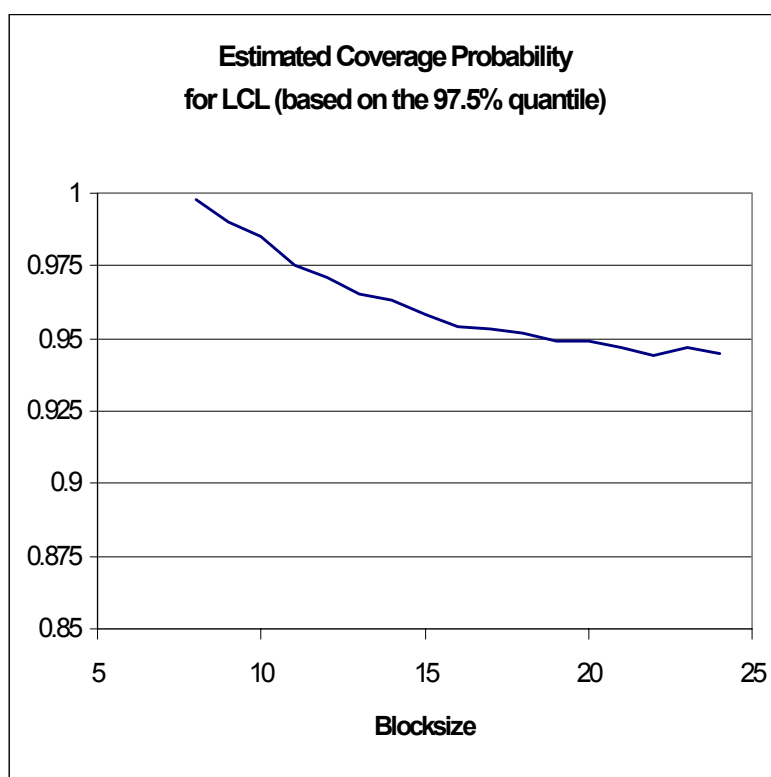


Figure 2: Estimated coverage probability as function of the blocksize b based on 1,000 residual-based bootstrap replications.

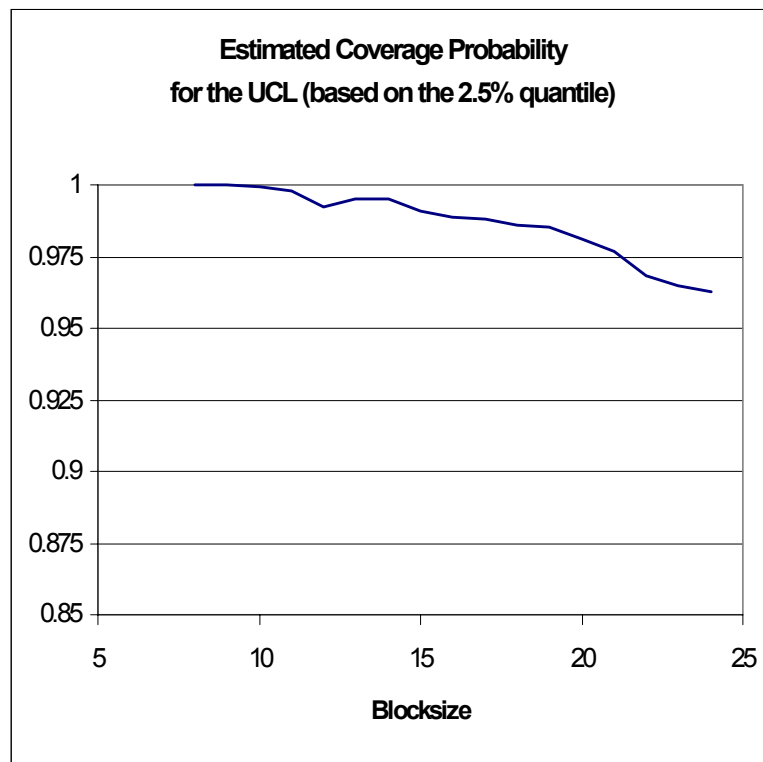


Table 1: Estimated coverage probabilities (in %) of nominal 95% equal-tailed and symmetric confidence intervals for the AR(1) parameter δ based on 1,000 replications and sample size $n = 100$. Block size is selected through minimising the volatility. The estimation model always includes an intercept and trend.

$$\beta = 0.0$$

$\delta \backslash \gamma$	Equal-tailed			Symmetric		
	0.6	0.9	0.95	0.6	0.9	0.95
0.6	76	77	80	90	91	92
0.9	80	80	79	92	90	89
0.95	82	81	80	90	86	86
1.0	84	83	81	82	78	75

$$\beta = 0.1$$

$\delta \backslash \gamma$	Equal-tailed			Symmetric		
	0.6	0.9	0.95	0.6	0.9	0.95
0.6	80	81	79	92	92	92
0.9	79	81	79	92	94	92
0.95	81	81	79	90	91	92
1.0	82	84	83	78	85	91

$$\beta = 0.4$$

$\delta \backslash \gamma$	Equal-tailed			Symmetric		
	0.6	0.9	0.95	0.6	0.9	0.95
0.6	78	80	80	90	92	94
0.9	79	81	79	93	96	96
0.95	83	79	78	93	96	96
1.0	86	82	79	87	96	97

Table 2: Estimated coverage probabilities (in %) of nominal 95% equal-tailed and symmetric confidence intervals for the AR(1) parameter δ based on 1,000 replications and a sample size $n = 100$. Block size is selected through the calibration algorithm, which is based on 250 bootstrap replications. The estimation model always includes an intercept and trend.

$$\beta = 0.0$$

$\delta \backslash \gamma$	Equal-tailed			Symmetric		
	0.6	0.9	0.95	0.6	0.9	0.95
0.6	94	94	94	94	95	96
0.9	94	94	93	95	93	92
0.95	94	93	93	93	91	90
1.0	93	93	91	86	85	83

$$\beta = 0.1$$

$\delta \backslash \gamma$	Equal-tailed			Symmetric		
	0.6	0.9	0.95	0.6	0.9	0.95
0.6	94	94	94	96	95	95
0.9	94	95	94	95	95	95
0.95	92	94	94	93	93	94
1.0	91	93	96	84	89	93

$$\beta = 0.4$$

$\delta \backslash \gamma$	Equal-tailed			Symmetric		
	0.6	0.9	0.95	0.6	0.9	0.95
0.6	94	93	95	95	96	96
0.9	93	96	95	94	97	96
0.95	94	94	95	95	96	96
1.0	93	94	95	89	96	98