SUBSAMPLING CONFIDENCE INTERVALS
FOR THE AUTOREGRESSIVE ROOT

by

JOSEPH P. ROMANO
Department of Statistics
Stanford University

and

MICHAEL WOLF
Department of Statistics
University of California, Los Angeles

May 1998

This research was supported in part by
National Science Foundation grant DMS-9704487

Department of Statistics
Sequoia Hall
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://www-stat.stanford.edu
Subsampling Confidence Intervals for the Autoregressive Root

Joseph P. Romano *  
Department of Statistics  
Stanford University  
Stanford, CA 94305

Michael Wolf  
Department of Statistics  
UCLA  
Los Angeles, CA 90095

May 1998

Abstract

In this paper, we propose a new method for constructing confidence intervals for the autoregressive parameter of an AR(1) model. Our method works when the parameter equals one, is close to one, or is far away from one and is therefore more general than previous procedures. The crux of the method is to recompute the OLS t-statistic for the AR(1) parameter on smaller blocks of the observed sequence, according to the subsampling approach of Politis and Romano (1994). Some simulation studies show good finite sample properties of our intervals.

KEY WORDS: Autoregressive Time Series, Local-To-Unity Asymptotics, Subsampling, Unit Roots.

*Joseph P. Romano, Phone: (650) 723-6326, Fax: (650) 725-8977, E-mail: romano@stat.stanford.edu.

We thank Graham Elliot and Guido Imbens for helpful discussions.
1 Introduction

This paper is concerned with making inference for $\rho$ in the simple AR(1) model

$$X_t = \mu + \rho X_{t-1} + \epsilon_t,$$  \hspace{1cm} (1)

where $\{\epsilon_t\}$ is a white noise innovation sequence and $|\rho| \leq 1$. It is well known that if $|\rho| < 1$, the sequence $\{X_t\}$ is strictly stationary with mean $\mu/(1-\rho)$. However, for $\rho = 1$ it is a random walk with drift. The parameter $\rho$ can be consistently estimated using OLS. Unfortunately, inference for $\rho$ is nontrivial, since the limiting distribution of the estimator depends on the underlying parameters $\rho$ and $\mu$ and on the distribution of $\{\epsilon_t\}$. If $|\rho| < 1$, the OLS estimator converges to a normal distribution at rate $n^{1/2}$. For $\rho = 1$ and $\mu = 0$, it converges to a nonstandard distribution at rate $n$. Finally, for $\rho = 1$ and $\mu \neq 0$, it converges to a normal distribution at rate $n^{3/2}$. To construct an asymptotic confidence interval for $\rho$ it therefore seems necessary to know which the correct model is. The discontinuity of the limiting distribution causes the bootstrap (based on resampling estimated residuals) to fail; see Basawa et al. (1991).

Note that the situation can be improved somewhat, but not remedied, by basing the inference on the OLS $t$-statistic. It will converge to a standard normal distribution in case $|\rho| < 1$ and the $\epsilon_t$ are i.i.d. or in case $\rho = 1$ and $\mu \neq 0$. On the other hand, if $|\rho| \leq 1$ and the $\epsilon_t$ are uncorrelated only, the limiting distribution is normal with mean zero but unknown variance. Finally, if $\rho = 1$ and $\mu = 0$, the limiting distribution is nonstandard; see Section 3 for details.

For this reason, much of the econometrics literature — under the heading of "unit root tests" — has been concerned with simply testing the null hypothesis of $\rho = 1$. One still needs to know whether $\mu = 0$ to derive the sampling distribution of the test statistic under the null. However, it turns that this problem can be avoided by including a time trend in the estimation process, that is, by using OLS to estimate $\rho$ from the following model

$$X_t = \mu + \delta t + \rho X_{t-1} + \epsilon_t.$$ \hspace{1cm} (2)

In this case it turns out that the estimator converges to a nonstandard, but fully known distribution at rate $n$ no matter what the value of $\mu$ is. The foundations of the unit root test literature were laid by Dickey and Fuller (1979) and Phillips and Perron (1988). A nice overview can be found in Chapter 17 of Hamilton (1994).

While testing for $\rho = 1$ is a worthwhile endeavor, it often leaves something to be desired. As Stock (1991) points out, "reporting only unit root tests and point estimates of the largest root is unsatisfying as a description of the data: this fails to convey information about the sampling uncertainty or, more precisely, the range of models (i.e., values of $\rho$) that are consistent with the observed data".
He proposes a method for finding confidence intervals for $\rho$ based on a slightly different model

\[ X_t = \alpha + \delta t + V_t, \quad V_t = \rho V_{t-1} + \epsilon_t. \] (3)

The method relies upon so-called local-to-unity asymptotics, which assume that $\rho$ shrinks to one as the sample size tends to infinity. More specifically,

\[ \rho = 1 + \frac{c}{n}, \] (4)

for some constant $c$. Typically, $c$ is thought to be less than or equal to zero, although the theory also works for positive $c$. It can be shown that under this model the OLS estimator converges at rate $n$ to a nonstandard distribution which depends on $c$ only (the distribution is a functional of an Ornstein-Uhlenbeck process). Therefore, one can test $H_0 : c = c_0$ for any arbitrary value of $c_0$ and a confidence interval for $c$ can be obtained as the collection of $c_0$ values not rejected by the test. Given the sample size $n$, one can convert the confidence interval for $c$ to one for $\rho$ using relation (4). Alternatively, confidence intervals can be based on the OLS $t$-statistic in a similar fashion. See Stock (1991) for details.

From simulation studies in Stock (1991) it appears that this method works well if $c = 0$ but gets worse the further $c$ is away from zero. This would imply that the method works well if $c = 0$ or if the sample size is small and $\rho$ is very close to one. For any $\rho \neq 1$, coverage probability of confidence intervals will gradually deteriorate as the sample size increases. Of course, this may be considered a philosophical rather than a practical problem, since in reality one is always faced with a fixed sample size.

It is the aim of this paper to provide a new way of finding confidence intervals for $\rho$. We propose a method that gives asymptotically correct results for all three cases of model (1) without knowing anything about the underlying parameters. Moreover, we consider $\rho$ to be fixed rather than a function of the sample size. Our intervals are based on the subsampling method.

The paper is organized as follows. In Section 2 we give a brief review of the subsampling method and provide some extensions of the previous theory needed for our application. Section 3 demonstrates that subsampling yields asymptotically correct confidence intervals for the AR(1) coefficient $\rho$. We discuss choosing the block size, an inherent model parameter of the subsampling method, in Section 4. Some finite sample simulation studies are reported in Section 5. Finally, the paper concludes with a summary in Section 6. All tables and figures appear after the references.
2 The Subsampling Method

2.1 The Basic Method

The subsampling methodology was introduced by Politis and Romano (1994) as an inference procedure that allows one to construct asymptotically valid confidence regions under very weak assumptions. For univariate data, their exposition encompasses the cases of i.i.d. and stationary observations. An extension to heteroscedastic observations was provided by Politis, Romano, and Wolf (1997). Those results could be used to make inference for the autoregressive parameter $\rho$ in the case $|\rho| < 1$. Clearly, we need a more general result that also works for the nonstationary case $\rho = 1$. Rather than presenting a custom-tailored answer pertaining to $\rho$, we shall derive a more general theorem for general univariate parameters $\theta$.

Suppose \{\ldots, X_{-1}, X_0, X_1, \ldots\} is a sequence of random variables taking values in an arbitrary sample space $S$, and defined on a common probability space. Denote the joint probability law governing the infinite sequence by $P$. The goal is to construct a confidence interval for some real-valued parameter $\theta = \theta(P)$, on the basis of observing $\{X_1, \ldots, X_n\}$. We assume the existence of a sensible estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$.

For time series data, the gist of the subsampling method is to recompute the statistic of interest on smaller blocks of the observed sequence $X_1, \ldots, X_n$. Define $\hat{\theta}_{b,a} = \hat{\theta}_b(X_a, \ldots, X_{a+b-1})$, the estimator of $\theta$ based on the subsample $X_a, \ldots, X_{a+b-1}$. In this notation $b$ is the block size and $a$ is the starting index of the smaller blocks. Note that $\hat{\theta}_{n,1} = \hat{\theta}_n$. Let $J_{b,a}(P)$ be the sampling distribution of $\tau_b (\hat{\theta}_{b,a} - \theta)$, where $\tau_b$ is an appropriate normalizing constant. Also define the corresponding cumulative distribution function:

$$J_{b,a}(x, P) = \text{Prob}_P \{\tau_b (\hat{\theta}_{b,a} - \theta(P)) \leq x\}. \quad (5)$$

A major assumption that we will need to construct asymptotically valid confidence intervals for $\theta$ is the following.

**Assumption 2.1** There exists a limiting law $J(P)$ such that

(i) $J_{n,1}(P)$ converges weakly to $J(P)$ as $n \to \infty$,

(ii) for every continuity point $x$ of $J(P)$, $\frac{1}{n-b+1} \sum_{a=1}^{n-b+1} J_{b,a}(x, P) \to J(x, P)$, for any sequences $n, b$ with $n, b \to \infty$ and $b/n \to 0$.

Condition (i) states that the estimator, properly normalized, has a limiting distribution. It is hard to conceive of any asymptotic theory free of such a requirement. Typically, much stronger assumptions are in force to ensure asymptotic normality.
Condition (ii) states that the distribution functions of the normalized estimator based on the subsamples will be on average close to the distribution function of the normalized estimator based on the entire sample, for large n.

**Remark 2.1** Note that condition (ii) follows trivially from condition (ii) if the process \{X_t\} is strictly stationary or, which is weaker, if the subsample statistics \( \tau_b(\hat{\theta}_{b,a} - \theta) \) are strictly stationary.

In order to describe our method, let \( Y_{b,a} \) be the block of size b of the consecutive data \{X_a, \ldots, X_{a+b-1}\}. Only a very weak assumption on b will be required. Typically, \( b/n \to 0 \) and \( b \to \infty \) as \( n \to \infty \). The subsampling approximation to \( J_n(x, P) \) we study is defined by

\[
L_n.b(x) = \frac{1}{n-b+1} \sum_{a=1}^{n-b+1} 1\{\tau_b(\hat{\theta}_{b,a} - \hat{\theta}_{n,1}) \leq x\}. 
\] (6)

The motivation behind the method is the following. For any \( a \), \( Y_{b,a} \) is a ‘true’ subsample of size b. Hence, the exact distribution of \( \tau_b(\hat{\theta}_{b,a} - \theta) \) is \( J_{b,a} \). If condition (ii) of Assumption 2.1 is satisfied, then the empirical distribution of the \( n - b + 1 \) values of \( \tau_b(\hat{\theta}_{b,a} - \theta) \) should serve as a good approximation to \( J_{n,1}(P) \), at least for large \( n \). Replacing \( \theta \) by \( \hat{\theta}_{n,1} \) is permissible because \( \tau_b(\hat{\theta}_{n,1} - \theta) \) is of order \( \tau_b/\tau_n \) in probability and we will assume that \( \tau_b/\tau_n \to 0 \).

Since we approximate \( J_{n,1}(x) \) by \( L_n.b(x) \), we want both to have the same limit, namely \( J(x, P) \). To ensure that \( L_n.b(x) \) converges to \( J(x, P) \) in probability it is necessary that the information in the \( n - b + 1 \) subsample statistics \( \tau_b(\hat{\theta}_{b,a} - \hat{\theta}_{n,1}) \) tend to infinity with the sample size \( n \). In previous theorems of Politis and Romano (1994) and Politis, Romano, and Wolf (1997) this followed from a weak dependence condition on the underlying sequence \{X_i\}, namely an \( \alpha \)-mixing condition. Roughly speaking, this means that \( X_i \) and \( X_j \) become independent as \( |i-j| \) tends to infinity. While this is true for an autoregressive series with \( |\rho| < 1 \), under some conditions on the innovation sequence, asymptotic independence is clearly violated for the nonstationary case \( \rho = 1 \). However, it turns out to be sufficient to impose a weak dependence condition on the random variables \( \tau_b(\hat{\theta}_{b,a} - \theta) \). To this end we have to give a formal definition of \( \alpha \)-mixing coefficients which generalizes the original notion due to Rosenblatt (1956).

**Definition 2.1** Given a random sequence \{\( Y_i \}\}, let \( \mathcal{F}_i \) be the \( \sigma \)-algebra generated by \( \{Y_i, Y_{i+1}, \ldots, Y_j\} \) and define the corresponding \( \alpha \)-mixing sequence by

\[
\alpha_Y(h) = \sup_i \sup_{A,B} |P(A \cap B) - P(A)P(B)|, 
\] (7)

where \( A \) and \( B \) vary over the \( \sigma \)-fields \( \mathcal{F}_i \) and \( \mathcal{F}_i^\infty \), respectively. The sequence \{\( Y_i \}\) is called \( \alpha \)-mixing or strong mixing if \( \alpha_Y(h) \to 0 \) as \( h \to \infty \).
The following theorem could be coined "a general asymptotic validity result under minimal conditions" and is an extension of previous results. Denote the mixing coefficients corresponding the \( n - b + 1 \) random variables \( \tau_n(\hat{\theta}_{b,a} - \theta) \) by \( \alpha_{n,b}(\cdot) \).

**Theorem 2.1** Assume Assumption 2.1 and that \( \tau_n/\tau_n \rightarrow 0, b/n \rightarrow 0 \) and \( b \rightarrow \infty \) as \( n \rightarrow \infty \). Also assume that \( n^{-1}\sum_{h=1}^{n} \alpha_{n,b}(h) \rightarrow 0 \). Let \( x \) be a continuity point of \( J(\cdot, P) \). Then

(i) \( L_{n,b}(x) \rightarrow J(x, P) \) in probability.

(ii) If \( J(\cdot, P) \) is continuous, then \( \sup_x |L_{n,b}(x) - J(x, P)| \rightarrow 0 \) in probability.

(iii) For \( \alpha \in (0, 1) \), let

\[
\begin{align*}
    c_{n,b,L}(1-\alpha) &= \inf\{x : L_{n,b}(x) \geq 1-\alpha\}, \\
    c_{n,b,U}(1-\alpha) &= \sup\{x : L_{n,b}(x) \leq 1-\alpha\}.
\end{align*}
\]

Correspondingly, define

\[
\begin{align*}
    c_{L}(1-\alpha, P) &= \inf\{x : J(x, P) \geq 1-\alpha\}, \\
    c_{U}(1-\alpha, P) &= \sup\{x : J(x, P) \leq 1-\alpha\}.
\end{align*}
\]

Let \( \{c_{n,b}(1-\alpha)\} \) be any sequence of random variables such that

\[
c_{n,b,L}(1-\alpha) \leq c_{n,b}(1-\alpha) \leq c_{n,b,U}(1-\alpha).
\]

In other words, \( c_{n,b}(1-\alpha) \) serves as an \( (1-\alpha) \) quantile of the subsampling distribution \( L_{n,b}(x) \).

If \( J(\cdot, P) \) is continuous at \( c_{L}(1-\alpha, P) \), then

\[
\text{Prob}_P\{\tau_n[\hat{\theta}_{n,1} - \theta(P)] \leq c_{n,b}(1-\alpha)\} \rightarrow 1-\alpha \quad \text{as} \quad n \rightarrow \infty.
\]

Thus, the asymptotic coverage probability under \( P \) of the interval

\[
I_1 = [\hat{\theta}_{n,1} - \tau_n^{-1}c_{n,b}(1-\alpha), \infty)
\]

is the nominal level \( 1-\alpha \).

**Proof of Theorem 2.1:** Without loss of generality we may think of \( b \) as a function of \( n \). Therefore we can reduce the notational burden by omitting the \( b \)-subscripts. For example, \( L_n(\cdot) \equiv L_{n,b}(\cdot) \), \( c_n(\alpha) \equiv c_{n,b}(\alpha) \), etc. To simplify the notation further introduce \( q \equiv q_n \equiv n - b + 1 \). Let

\[
U_n(x) = \frac{1}{q} \sum_{a=1}^{q} 1\{\tau_a[\hat{\theta}_{a,a} - \theta(P)] \leq x\}.
\]
To prove (i), it suffices to show that \( U_n(x) \) converges in probability to \( J(x, P) \) for every continuity point \( x \) of \( J(\cdot, P) \). This can be seen by noting that

\[
L_n(x) = \frac{1}{q} \sum_{a=1}^{q} 1\{\tau_b[\hat{\theta}_{b,a} - \theta(P)] + \tau_b[\theta(P) - \hat{\theta}_{n,1}] \leq x\},
\]

so that for every \( \epsilon > 0 \),

\[
U_n(x - \epsilon)1\{E_n\} \leq L_n(x)1\{E_n\} \leq U_n(x + \epsilon),
\]

where \( 1\{E_n\} \) is the indicator of the event \( E_n = \{\tau_b[\theta(P) - \hat{\theta}_{n,1}] \leq \epsilon\} \). But, the event \( E_n \) has probability tending to one. So, with probability tending to one,

\[
U_n(x - \epsilon) \leq L_n(x) \leq U_n(x + \epsilon).
\]

Thus, if \( x + \epsilon \) and \( x - \epsilon \) are continuity points of \( J(\cdot, P) \), then \( U_n(x \pm \epsilon) \to J(x \pm \epsilon, P) \) in probability implies

\[
J(x - \epsilon, P) - \epsilon \leq L_n(x) \leq J(x + \epsilon, P) + \epsilon
\]

with probability tending to one. Now, let \( \epsilon \to 0 \) such that \( x \pm \epsilon \) are continuity points of \( J(\cdot, P) \) to conclude that \( L_n(x) \to J(x, P) \) in probability as well. Therefore, we may restrict our attention to \( U_n(x) \).

Since \( E(U_n(x)) = \frac{1}{q} \sum_{a=1}^{q} J_{b,a}(x) \), the proof of (i) reduces by Assumption 2.1 to showing that \( \text{Var}(U_n(x)) \) tends to zero. Define

\[
I_{b,a} = 1\{\tau_b[\hat{\theta}_{b,a} - \theta(P)] \leq x\}, \quad a = 1, \ldots, q,
\]

\[
s_{q,h} = \frac{1}{q} \sum_{a=1}^{q-h} \text{Cov}(I_{b,a}, I_{b,a+h}).
\]

Due to a standard mixing inequality for bounded random variables, \( |\text{Cov}(I_{b,a}, I_{b,a+h})| \leq 4\alpha_{n,b}(h) \) and therefore,

\[
\text{Var}(U_n(x)) = \frac{1}{q} \left( s_{q,0} + 2 \sum_{h=1}^{q-1} s_{q,h} \right)
\]

\[
\leq \frac{1}{q} \left( 1 + 2 \sum_{h=1}^{q-1} \alpha_{n,b}(h) \right) \to 0.
\]

This completes the proof of (i).

To prove (ii), given any subsequence \( \{n_k\} \), one can extract a further subsequence \( \{n_{k_j}\} \) such that \( L_{n_{k_j}}(x) \to J(x, P) \) for all \( x \) in some countable dense set of the real line, almost surely. It then follows that, on a set of probability one, \( L_{n_{k_j}}(x) \) tends weakly to \( J(x, P) \). By the continuity of \( J(\cdot, P) \) this convergence is uniform by Polya's theorem.
The proof of (iii) is very similar to the proof of Theorem 1 of Beran (1984) given our result (i) and is omitted. ■

The interval \( I_1 \) in (iii) corresponds to a one-sided hybrid percentile interval in the bootstrap literature (e.g., Hall, 1992). A two-sided equal-tailed confidence interval can be obtained by forming the intersection of two one-sided intervals. The two-sided analogue of \( I_1 \) is

\[
I_2 = [\hat{\theta}_n - \tau_n^{-1} c_{n,b}(1 - \alpha/2), \hat{\theta}_n - \tau_n^{-1} c_{n,b}(\alpha/2)].
\]

\( I_2 \) is called equal-tailed because it has approximately equal probability in each tail:

\[
Prob_P\{\theta < \hat{\theta}_n - \tau_n^{-1} c_{n,b}(1 - \alpha/2)\} \doteq Prob_P\{\theta > \hat{\theta}_n - \tau_n^{-1} c_{n,b}(\alpha/2)\} \doteq \alpha/2.
\]

As an alternative approach, two-sided symmetric confidence intervals can be constructed. A two-sided symmetric confidence interval is given by \([\hat{\theta}_n - \hat{c}, \hat{\theta}_n + \hat{c}]\), where \( \hat{c} \) is chosen so that \( Prob_P\{|\hat{\theta}_n - \theta| > \hat{c}\} = \alpha \). Hall (1988) showed that symmetric bootstrap confidence intervals enjoy enhanced coverage and, even in asymmetric circumstances, can be shorter than equal-tailed confidence intervals. To construct two-sided symmetric subsampling intervals in practice we estimate the two-sided distribution function

\[
J_{n,1,1,1}(x, P) = Prob_P\{\tau_n |\hat{\theta}_n - \theta| \leq x\}.
\]

The subsampling approximation to \( J_{n,1,1,1}(x, P) \) is defined by

\[
L_{n,b,1,1}(x) = \frac{1}{n - b + 1} \sum_{a=1}^{n-b+1} 1\{\tau_b |\hat{\theta}_{b,a} - \hat{\theta}_n| \leq x\}.
\]

From Theorem 2.1 we can immediately follow the asymptotic validity of two-sided symmetric subsampling intervals.

**Corollary 2.1** Make the same assumptions as in Theorem 2.1 and denote by \( J_{1,1,1}(, P) \), \( c_{L,1,1} \) and \( c_{n,b,1,1} \) the obvious. Let \( x \) be a continuity point of \( J_{1,1,1}(, P) \). Then

(i) \( L_{n,b,1,1}(x) \rightarrow J_{1,1,1}(x, P) \) in probability.

(ii) If \( J_{1,1,1}(, P) \) is continuous, then \( sup_x |L_{n,b,1,1}(x) - J_{1,1,1}(x, P)| \rightarrow 0 \) in probability.

(iii) If \( J_{1,1,1}(, P) \) is continuous at \( c_{L,1,1}(1 - \alpha, P) \), then

\[
Prob_P\{\tau_n |\hat{\theta}_n - \theta| \leq c_{n,b,1,1}(1 - \alpha)\} \rightarrow 1 - \alpha \quad \text{as} \; n \rightarrow \infty.
\]

Thus, the asymptotic coverage probability under \( P \) of the interval

\[
I_{SYM} = [\hat{\theta}_n - \tau_n^{-1} c_{n,b,1,1}(1 - \alpha), \hat{\theta}_n + \tau_n^{-1} c_{n,b,1,1}(1 - \alpha)]
\]

is the nominal level \( 1 - \alpha \).

**Proof of Corollary 2.1:** The proof follows immediately from Theorem 2.1 and the continuous mapping theorem. ■
2.2 Subsampling Studentized Statistics

The application of Theorem 2.1 or Corollary 2.1 requires knowledge of the rate of convergence $\tau_n$. In standard cases this is simply $n^{1/2}$. However, nonstandard cases exist and we are interested in the unit root case where the rate of convergence can be given by $n$ or even $n^{3/2}$ (see Section 3). Therefore, subsampling inference for the autoregressive root would require knowledge about the underlying model. But this is exactly the problem that we are trying to solve! Fortunately, we will be able to get around this dilemma by basing inference on a studentized statistic instead. This requires an extension of the theory so far. We stated the previous results nevertheless, since they are interesting in their own right but also since they provide the basis for the proof of the theorem handling the studentized case.

The focus is now on a studentized root $\tau_n(\hat{\theta}_n - \theta)/\hat{\sigma}_n$, where $\hat{\sigma}_n$ is some nonnegative estimate of scale. Define $J_{b,a}^*(P)$ to be the sampling distribution of $\tau_b(\hat{\theta}_{b,a} - \theta)/\hat{\sigma}_{b,a}$ based on the subsample $X_{a}, \ldots, X_{a+b-1}$. Also define the corresponding cumulative distribution function

$$J_{b,a}^*(x,P) = \text{Prob}_P\{\tau_b(\hat{\theta}_{b,a} - \theta)/\hat{\sigma}_{b,a} \leq x\}.$$  

Subsampling for scaled or studentized statistics in the context of i.i.d. data has previously been considered in Politis and Romano (1993) in case where $\hat{\sigma}_n$ converges to a positive constant in probability and by Romano and Wolf (1997) in case where $\hat{\sigma}_n$ (up to a scale sequence) converges weakly. For our purposes we need a result that covers dependent data.

The subsampling method is modified to the studentized case in the obvious way. Analogous to (6) define

$$L_{n,b}^*(x) = \frac{1}{n - b + 1} \sum_{a=1}^{n-b+1} 1\{\tau_b(\hat{\theta}_{b,a} - \hat{\theta}_n)/\hat{\sigma}_{b,a} \leq x\}.  \quad (10)$$

$L_{n,b}^*(x)$ then represents the subsampling approximation to $J_{n,1}^*(x)$.

The essential assumption needed to construct asymptotically valid confidence regions for $\theta$ now becomes more involved than for the non-studentized case.

Assumption 2.2

(i) $J_n^*(P)$ converges weakly to a nondegenerate limit law $J^*(P)$ as $n \to \infty$. In addition, $a_n(\hat{\theta}_n - \theta)$ converges weakly to $V$, and $d_n\hat{\sigma}_n$ converges weakly to $W$. Here, $V$ and $W$ are two random variables, with distributions $V(P)$ and $W(P)$. $W(P)$ does not have positive mass at zero. Of course, $\tau_n = a_n/d_n$.

(ii) For every continuity point $x$ of $J^*(P)$, $\frac{1}{n - b + 1} \sum_{a=1}^{n-b+1} J_{b,a}^*(x,P) \to J^*(x,P)$, for any sequences $n, b$ with $n, b \to \infty$ and $b/n \to 0$. 

8
(iii) for every continuity point \( x \) of \( \mathbb{W}(P) \), \( \frac{1}{n^{-b+1}} \sum_{a=1}^{n^{-b+1}} W_{b,a}(x, P) \to W(x, P) \), for any sequences \( n, b \) with \( n, b \to \infty \) and \( b/n \to 0 \). Here \( W_{b,a}(x, P) = \text{Prob}_P\{d_b \hat{\sigma}_{b,a} \leq x\} \).

Remark 2.2 Note that condition (ii) follows trivially from condition (ii) if the process \( \{X_t\} \) is strictly stationary or, which is weaker, if the subsample statistics \( \tau_b(\hat{\theta}_{b,a} - \theta)/\hat{\sigma}_{b,a} \) are strictly stationary.

Denote the mixing coefficients corresponding the \( n-b+1 \) random vectors \( (\tau_b(\hat{\theta}_{b,a} - \theta)/\hat{\sigma}_{b,a}) \) by \( \alpha_n(\cdot) \).

Theorem 2.2 Assume Assumption 2.2, \( a_b/a_n \to 0, \tau_b/\tau_n \to 0, b/n \to 0 \) and \( b \to \infty \) as \( n \to \infty \). Also assume that \( n^{-1} \sum_{n=1}^{n} \alpha_n(\cdot) \to 0 \). Let \( x \) be a continuity point of \( J^*(\cdot, P) \). Then

(i) \( L^*_{n,b}(x) \to J^*(x, P) \) in probability.

(ii) If \( J^*(\cdot, P) \) is continuous, then \( \sup_x |L^*_{n,b} - J^*(x, P)| \to 0 \) in probability.

(iii) For \( \alpha \in (0,1) \), let \( c^*_{n,b}(1-\alpha) = \inf\{x : L^*_{n,b}(x) \geq 1-\alpha\} \). Correspondingly, define \( c^*(1-\alpha, P) = \inf\{x : J^*(x, P) \geq 1-\alpha\} \). If \( J^*(\cdot, P) \) is continuous at \( c^*(1-\alpha, P) \) then

\[
\text{Prob}_P\{\tau(\hat{\theta}_n - \theta)/\hat{\sigma}_n \leq c^*_{n,b}(1-\alpha)\} \to 1-\alpha
\]

as \( n \to \infty \). Thus, the asymptotic coverage probability under \( P \) of the interval \( I_1^* = [\hat{\theta}_n - \tau_n^{-1} c^*_{n,b}(1-\alpha), \infty) \) is the nominal level \( 1-\alpha \).

Proof of Theorem 2.2: Again let \( q = n-b+1 \). To prove (i) note that

\[
L^*_{n,b}(x) = \frac{1}{q} \sum_{a=1}^{q} 1\{\tau_b(\hat{\theta}_{b,a} - \hat{\theta}_n)/\hat{\sigma}_{b,a} \leq x\}
\]

\[
= \frac{1}{q} \sum_{a=1}^{q} 1\{\tau_b(\hat{\theta}_{b,a} - \theta)/\hat{\sigma}_{b,a} \leq x + \tau_b(\hat{\theta}_n - \theta)/\hat{\sigma}_{b,a}\}.
\]

We want to show that the terms \( \tau_b(\hat{\theta}_n - \theta)/\hat{\sigma}_{b,a} \) are negligible in the last equation. To this end, for \( t > 0 \), let

\[
R_n(t) = \frac{1}{q} \sum_{a=1}^{q} 1\{\tau_b(\hat{\theta}_n - \theta)/\hat{\sigma}_{b,a} \leq t\}
\]

\[
= \frac{1}{q} \sum_{a=1}^{q} 1\{d_b \hat{\sigma}_{b,a} \geq a_b(\hat{\theta}_n - \theta)/t\}.
\]
Here, we are assuming without loss of generality that both the sequences \( a_n \) and \( b_n \) are nonnegative. By Assumption 2.2 and \( a_n/a_n \to 0 \), we have for any \( \delta > 0 \) that \( a_b(\hat{\theta}_n - \theta) \leq \delta \) with probability tending to one. Therefore, with probability tending to one

\[
R_n(t) \geq \frac{1}{q} \sum_{a=1}^{q} 1\{d_b\hat{\sigma}_{b,a} \geq \delta / t\}.
\]

We need to consider the case \( t > 0 \) only, as the scale estimates \( \hat{\sigma}_{b,a} \) are nonnegative. Due to the usual subsampling argument of Theorem 2.1, \( \frac{1}{q} \sum_{a=1}^{q} 1\{d_b\hat{\sigma}_{b,a} \geq \delta / t\} \) converges in probability to \( 1 - W(\delta / t, P) \), as long as \( \delta / t \) is a continuity point of \( W(P) \); note that we do not require \( d_b/d_n \to 0 \) here since the subsample statistics are of the form \( d_b\hat{\sigma}_{b,a} \) rather than \( d_b(\hat{\sigma}_{b,a} - \hat{\sigma}_n) \).

Hence, we can make sure that \( R_n(t) \) is arbitrarily close to one by choosing \( \delta \) small enough; remember we assume that \( W(P) \) does not have positive mass at zero. In other words, for any \( t > 0 \), we have \( R_n(t) \to 1 \) in probability. Let us now rewrite (11) in the following way

\[
L_{n,b}^*(x) = \frac{1}{q} \sum_{a=1}^{q} 1\{\tau_b(\hat{\theta}_{b,a} - \theta) / \hat{\sigma}_{b,a} \leq x + \tau_b(\hat{\theta}_n - \theta) / \hat{\sigma}_{b,a}\} \\
\leq \frac{1}{q} \sum_{a=1}^{q} 1\{\tau_b(\hat{\theta}_{b,a} - \theta) / \hat{\sigma}_{b,a} \leq x + t\} + (1 - R_n(t)),
\]

for any positive number \( t \). The last inequality follows because the \( a \)-th term in (11) is less than or equal to

\[
1\{\tau_b(\hat{\theta}_{b,a} - \theta) / \hat{\sigma}_{b,a} \leq x + t\} + 1\{\tau_b(\hat{\theta}_n - \theta) / \hat{\sigma}_{b,a} > t\};
\]

then, sum over all \( a \). We have seen that \( (1 - R_n(t)) \to 0 \) in probability and hence by a standard subsampling argument again we get, for any \( \epsilon > 0 \), \( L_{n,b}^*(x) \leq J^*(x + t, P) + \epsilon \) with probability tending to one, provided that \( x + t \) is a continuity point of \( J^*(\cdot, P) \). Letting \( t \) tend to zero shows that \( L_{n,b}^*(x) \leq J^*(x, P) + \epsilon \) with probability tending to one. A similar argument leads to \( L_{n,b}^*(x) \geq J^*(x, P) - \epsilon \) with probability tending to one. Since \( \epsilon \) is arbitrary, this implies \( L_{n,b}^* \to J^*(x, P) \) in probability, and thus we have proved (i).

The proofs of (ii) and (iii) given (i) are very similar to the proofs of (ii) and (iii) given (i) in Theorem 2.1 and thus are omitted. ■

The issue of symmetric confidence intervals as discussed by Corollary 2.1 for subsampling non-studentized statistics applies here as well. Let \( J_{n,1,1,1,1}(P) \) be the sampling distribution of \( \tau_n \big| \hat{\theta}_n - \theta \big| / \hat{\sigma}_n \). Define

\[
L_{n,b,1,1}(x) = \frac{1}{n - b + 1} \sum_{a=1}^{n-b+1} 1\{\tau_b \big| \hat{\theta}_{b,a} - \hat{\theta}_n \big| / \hat{\sigma}_{b,a} \leq x\}.
\]

\( L_{n,b,1,1}(x) \) then represents the subsampling approximation to \( J_{n,1,1,1,1}^*(x) \).
Corollary 2.2 Assume Assumption 2.2, \( a_n/a_n \to 0 \), \( \tau_n/\tau_n \to 0 \), \( b_n/b_n \to 0 \) and \( b \to \infty \) as \( n \to \infty \). Also assume that \( n^{-1} \sum_{h=1}^{n} \alpha_{n,b}(h) \to 0 \). Let \( x \) be a continuity point of \( J^*_{\| \cdot \|}(x, P) \), where \( J^*_{\| \cdot \|}(P) \) is the law of \( |Y| \) for \( Y \sim J^*(P) \). Then

(i) \( L^*_{n,b\| \cdot \|}(x) \to J^*_{\| \cdot \|}(x, P) \) in probability.

(ii) If \( J^*_{\| \cdot \|}(x, P) \) is continuous, then \( \sup_x |L^*_{n,b\| \cdot \|}(x, P) - J^*_{\| \cdot \|}(x, P)| \to 0 \) in probability.

(iii) For \( \alpha \in (0,1) \), let \( c^*_{n,b\| \cdot \|}(1-\alpha) = \inf\{ x : L^*_{n,b\| \cdot \|}(x) \geq 1-\alpha \} \). Correspondingly, define \( c^*_{\| \cdot \|}(1-\alpha, P) = \inf\{ x : J^*_{\| \cdot \|}(x, P) \geq 1-\alpha \} \). If \( J^*_{\| \cdot \|}(x, P) \) is continuous at \( c^*_{\| \cdot \|}(1-\alpha, P) \) then

\[
Prob_P\{ \tau_n |\hat{\theta}_n - \theta| / \tilde{\sigma}_n \leq c^*_{n,b\| \cdot \|}(1-\alpha) \} \to 1 - \alpha
\]

as \( n \to \infty \). Thus, the asymptotic coverage probability under \( P \) of the interval

\[
I_{\text{SYM}} = [\hat{\theta}_n - \tilde{\sigma}_n \tau_n^{-1} c^*_{n,b\| \cdot \|}(1-\alpha), \hat{\theta}_n + \tilde{\sigma}_n \tau_n^{-1} c^*_{n,b\| \cdot \|}(1-\alpha)]
\]

is the nominal level \( 1 - \alpha \).

Proof of Corollary 2.2: The proof follows immediately from Theorem 2.2 and the continuous mapping theorem.

3 Subsampling Inference for the Autoregressive Root

The goal of this section is to show that the subsampling method for studentized statistics as outlined in the previous section can be used to make inference for the autoregressive root. We observe a sample \( X_1, \ldots, X_n \) from the our AR(1) model including a drift:

\[
X_t = \mu + \rho X_{t-1} + \epsilon_t,
\]

where \( \{\epsilon_t\} \) is a white noise innovation sequence with mean zero and variance \( \sigma^2_\epsilon \). Hence, we assume that the \( \epsilon_t \) are strictly stationary and uncorrelated, but possibly dependent. Note that the assumption of strict stationarity is mainly made for convenience of the proofs but could be relaxed considerably; see Remark 3.4. Denote the \( \alpha \)-mixing coefficients corresponding to \( \{\epsilon_t\} \) by \( \alpha_\epsilon(\cdot) \). Our estimator for \( \rho \) is obtained by OLS regression of \( X_t \) on \( X_{t-1} \), including an intercept in the regression. To facilitate the notation, let \( \bar{X}_{i,j} = (j - i + 1)^{-1} \sum_{t=i}^{j} X_t \).

\[
\hat{\rho}_n = \frac{\sum_{t=i}^{n-1}(X_{t+1} - \bar{X}_{2,n})(X_t - \bar{X}_{1,n-1})}{\sum_{t=i}^{n-1}(X_t - \bar{X}_{1,n-1})^2}, \quad \hat{\mu}_n = \bar{X}_{2,n} - \hat{\rho}_n \bar{X}_{1,n}.
\]

The corresponding studentized statistic is based on the OLS estimator for \( SD(\hat{\rho}_n) \).

\[
\tilde{S}_D(\hat{\rho}_n) = \frac{s_n}{[\sum_{t=1}^{n-1}(X_t - \bar{X}_{1,n-1})^2]^{1/2}}, \quad s^2_n = \frac{1}{n-3} \sum_{t=1}^{n-1}(X_{t+1} - \bar{X}_{2,n} + \hat{\rho}_n \bar{X}_{1,n-1} - \hat{\rho}_n X_t)^2.
\]
To fit this application in the framework of the theory of Subsection 2.2, we are going to use

\[ \hat{\sigma}_n = n^{1/2} SD(\hat{\rho}_n), \] (17)

so the studentized statistic and the corresponding subsample statistics are of the form

\[ n^{1/2} \frac{(\hat{\rho}_n - \rho)}{\hat{\sigma}_n}, \quad b^{1/2} \frac{(\hat{\rho}_{b,a} - \hat{\rho}_n)}{\hat{\sigma}_{b,a}} \] (18)

Here, \( \hat{\rho}_{b,a} \) and \( \hat{\sigma}_{b,a} \) are \( \hat{\rho}_b \) and \( \hat{\sigma}_b \) computed on the subsample \( X_a, \ldots, X_{a+b-1} \).

The following proposition shows that the ideas of Subsection 2.2 can be used to construct asymptotic confidence intervals for \( \rho \).

**Proposition 3.1** Assume \( \{X_t\} \) is a sequence of random variables according to model (14). Assume that for some \( \delta > 0 \), \( E |\epsilon_t|^{4+\delta} < \infty \), \( \sum_{h=1}^{\infty} \alpha_t^{(2+\delta)}(h) < \infty \), and, if \( |\rho| < 1 \), \( \sum_{h=1}^{\infty} \alpha_X^{(2+\delta)}(h) < \infty \). Take \( \hat{\rho}_n \) and \( \hat{\sigma}_n \) as defined in (15) and (17). Also, let \( \tau_n = n^{1/2} \) and assume that \( b \to \infty \) and \( b/n \to 0 \) as \( n \to \infty \).

Then the conclusions of Theorem 2.2 and Corollary 2.2 hold.

**Proof of Proposition 3.1:** For the proof we have to show that Assumption 2.2 holds, that the subsample statistics satisfy the mixing condition \( n^{-1} \sum_{h=1}^{\infty} \alpha_{n,h}^{*}(h) \to 0 \), and that \( a_b/a_n \to 0 \). We start with Assumption 2.2 and \( a_b/a_n \to 0 \). First consider the stationary case \( |\rho| < 1 \). It is well known (e.g., Hamilton, 1994) that if the \( \epsilon_t \) are i.i.d., then

\[ n^{1/2}(\hat{\rho}_n - \rho) \overset{L}{\to} N(0, 1 - \rho^2), \]

\[ \hat{\sigma}_n^2 \overset{P}{\to} 1 - \rho^2, \]

and

\[ n^{1/2} \frac{(\hat{\rho}_n - \rho)}{\hat{\sigma}_n} \overset{L}{\to} N(0, 1). \]

Here, \( \overset{L}{\to} \) denotes convergence in distribution and \( \overset{P}{\to} \) denotes convergence in probability. Actually, these results are frequently used to make asymptotic inference for \( \rho \) in the stationary case. However, that inference can be arbitrarily misleading if the \( \epsilon_t \) are uncorrelated but dependent, as was pointed out by Romano and Thombs (1996). Under the conditions of Proposition 3.1, they proved the following result

\[ n^{1/2}(\hat{\rho}_n - \rho) \overset{L}{\to} N(0, \xi^2). \]

Here, the limiting variance is given by

\[ \xi^2 = \lim_n Var(\hat{\rho}_n) = Var^{-2}(X_1)[c_{2,2} - 2 \rho c_{1,2} + \rho^2 c_{1,1}], \]

(19)
where \( c_{i,j} \) is defined as follows

\[
c_{i+1,j+1} = \sum_{d=-\infty}^{\infty} \text{Cov} \left( X_0 X_i, X_d X_{d+j} \right).
\]  

(20)

If the \( \epsilon_t \) are i.i.d., then \( \xi^2 = 1 - \rho^2 \). But as is clear from equations (19) and (20), in the dependent case this can change to any positive value, as a function of the underlying dependence structure; see Romano and Thombs (1996) for some explicit examples. On the other hand, the relation

\[
\delta_n^2 \overset{P}{\to} 1 - \rho^2
\]

remains unchanged so that for the studentized statistic we get

\[
n^{1/2} \frac{\hat{\rho}_n - \rho}{\delta_n} \overset{\mathcal{D}}{\to} N(0, \frac{\xi^2}{1 - \rho^2}).
\]

(22)

Since the limiting variance can be very different from one, basing inference for \( \rho \) on limiting standard normality of the OLS \( t \)-statistic can be completely misleading if the \( \epsilon_t \) are uncorrelated only. It is the point of this proof to show that subsampling, on the other hand, will give correct answers. Due to stationarity it follows that the distributions of \( b^{1/2}(\hat{\rho}_{b,a} - \rho)/\delta_{b,a} \) and \( \delta_{b,a} \) are independent of \( b \). And from (22) and (21) we see that the respective limiting distributions are \( N(0, \frac{\xi^2}{1 - \rho^2}) \) and point mass at \( 1 - \rho^2 \). Therefore all three conditions of Assumption 2.2 are met. Since \( \tau_n = n^{1/2} \) and \( d_n \equiv 1, a_n = n^{1/2} \). Thus, \( a_{b}/a_{n} \to 0 \) and \( a_{b}/a_{n} \to 0 \) trivially follow from \( b/n \to 0 \).

Next we consider the nonstationary case \( \rho = 1 \). Surprisingly at first sight, the asymptotics in this case strongly depend upon whether \( \mu = 0 \). We discuss the instance \( \mu = 0 \) first. In this case it is known (e.g., Hamilton, 1994) that

\[
n(\hat{\rho}_n - 1) \overset{\mathcal{D}}{\to} \frac{1}{2} \{ [W(1)]^2 - 1 \} - W(1) \int_0^1 W(r)dr
\]

\[
\int_0^1 W(r)^2 dr - \left[ \int_0^1 W(r)dr \right]^2,
\]

(23)

\[
n^{1/2} \delta_n \overset{\mathcal{D}}{\to} \frac{1}{\left\{ \int_0^1 W(r)^2 dr - \left[ \int_0^1 W(r)dr \right]^2 \right\}^{1/2}},
\]

(24)

and

\[
n^{1/2} \frac{\hat{\rho}_n - 1}{\delta_n} \overset{\mathcal{D}}{\to} \frac{1}{2} \{ [W(1)]^2 - 1 \} - W(1) \int_0^1 W(r)dr
\]

\[
\left\{ \int_0^1 W(r)^2 dr - \left[ \int_0^1 W(r)dr \right]^2 \right\}^{1/2},
\]

(25)

where \( W(\cdot) \) denotes a Brownian Motion. These result follows from a functional central limit theorem for the \( \{ X_t \} \) process and were first proved by Dickey and Fuller (1979) for i.i.d. innovations \( \epsilon_t \). Phillips and Perron (1988) extended them to uncorrelated observations. A summary of these results can be found in Hamilton (1994). It is easy to see that the distributions of
are independent of \( b \). The respective limiting distributions are given by (25) and (24); note that the distribution defined by (24) does not have positive mass at zero. Therefore all three conditions of Assumption 2.2 are met. Furthermore, we have \( \tau_n = n^{1/2}, \; d_n = n^{1/2}, \) and \( a_n = n \). Thus, \( \tau_b / \tau_n \to 0 \) and \( a_b / a_n \to 0 \) trivially are implied by \( b / n \to 0 \).

Things change dramatically for \( \mu \neq 0 \). The asymptotic results are

\[
n^{3/2}(\hat{\rho}_n - 1) \overset{D}{\rightarrow} N(0, \sigma_e^2 \mu^2 / 12), \tag{26}
\]

\[
n \hat{\sigma}_n \overset{P}{\rightarrow} \sigma_e \mu / \sqrt{12}, \tag{27}
\]

and

\[
n^{1/2}(\hat{\rho}_n - 1) / \hat{\sigma}_n \overset{D}{\rightarrow} N(0, 1). \tag{28}
\]

The proof follows from Section 2 of West (1988). The reason for the discrepancy to the previous case is that for \( \mu \neq 0 \) the regressor \( X_{t-1} \) is asymptotically dominated by the time trend \( \mu (t - 1) \). In large samples, it is as if the explanatory variable \( X_{t-1} \) were replaced by the time trend \( \mu (t - 1) \). Superconsistency of \( \hat{\rho}_n \) at rate \( n^{3/2} \) and asymptotic standard normality of the studentized statistic ensue. Note that in contrast to the stationary case \( |\rho| < 1 \) the limiting variance of the \( t \)-statistic is always one and not affected by the dependence structure of the \( \epsilon_t \). See West (1988) for details or Hamilton (1994) for an analogous discussion concerning i.i.d. innovations. It is easy to see that the distributions of \( b^{1/2}(\hat{\rho}_{b,a} - \rho) / \hat{\sigma}_{b,a} \) and \( b \hat{\sigma}_{b,a} \) are independent of \( b \), with the respective limiting distributions being \( N(0, 1) \) and point mass at \( \sigma_e \mu / \sqrt{12} \). Therefore all three conditions of Assumption 2.2 are met. Furthermore, we have \( \tau_n = n^{1/2}, \; d_n = n, \) and \( a_n = n^{3/2} \). Thus, \( \tau_b / \tau_n \to 0 \) and \( a_b / a_n \to 0 \) trivially are implied by \( b / n \to 0 \).

Finally, we show that the subsample statistics satisfy the necessary mixing condition. Note that for the stationary case \( |\rho| < 1 \) this condition immediately follows from the mixing condition of the \( \{X_t\} \) process. However, for \( \rho = 1 \) this is no longer true. For example, if the innovation sequence \( \{\epsilon_t\} \) is i.i.d. with unit variance, then \( \text{Cov}(X_t, X_{t+j}) = i \) which does not vanish as \( j \to \infty \). For this reason we have to verify the weak dependence structure of the subsampling statistics from first principles in the case \( \rho = 1 \). To this end we claim that in this case \( \hat{\rho}_{b,a} \) and \( \hat{\sigma}_{b,a} \) are functions of \( \mu \) and \( \epsilon_{\alpha_1}, \ldots, \epsilon_{\alpha + a - 1} \) only. From this claim and with \( \mu \) being a constant it follows that \( \alpha^*_{n,b}(b + h) \leq \alpha^*_t(h) \) and therefore that

\[
\frac{1}{n} \sum_{h=1}^{n} \alpha^*_{n,b} \leq \frac{b}{n} + \frac{1}{n} \sum_{h=1}^{n} \alpha^*_t(h).
\]

Both terms on the right hand side converge to 0 by our assumptions. To demonstrate our
claim, note that for \( i \leq k \leq j \)

\[
X_k - \bar{X}_{i,j} = X_k - \frac{1}{j - i + 1} \sum_{t=i}^{j} X_t
\]

\[
= X_i + (k - i)\mu + \sum_{t=i+1}^{k} \epsilon_t - \frac{1}{j - i + 1} \sum_{t=i}^{j} (X_t + (t - i)\mu) + \sum_{t=i+1}^{t} \epsilon_t
\]

\[
= (k - i + \frac{j - i}{2})\mu + \sum_{t=i+1}^{k} \epsilon_t - \frac{1}{j - i + 1} \sum_{t=i}^{j} (t - i + 1)\epsilon_t.
\]

(29)

Hence, for \( i \leq k \leq j \), \( X_k - X_{i,j} \) is a function of \( \mu \) and \( \epsilon_i, \ldots, \epsilon_j \) only. From this observation and equation (15) it follows immediately that \( \hat{\rho}_{b,a} \) is a function of \( \mu \) and \( \epsilon_a, \ldots, \epsilon_{a+b-1} \) only. To see that the same is true for \( \hat{\sigma}_{b,a} \), we conclude from equations (15) to (17) that we are left to show that \( \hat{s}_{b,a}^2 \) is a function of \( \mu \) and \( \epsilon_a, \ldots, \epsilon_{a+b-1} \) only. This is implied by

\[
\hat{s}_{b,a}^2 = \frac{1}{b - 3} \sum_{t=a}^{b+a-2} (X_{t+1} - \bar{X}_{2,b+a-1} + \hat{\rho}_{b,a} \bar{X}_{a,b+a-2} - \hat{\rho}_{b,a} X_t)^2
\]

\[
= \frac{1}{b - 3} \sum_{t=a}^{b+a-2} (X_{t+1} - \bar{X}_{2,b+a-1} + \hat{\rho}_{b,a} (X_t - \bar{X}_{a,b+a-2}))^2,
\]

relationship (29), and \( \hat{\rho}_{b,a} \) being a function of \( \mu \) and \( \epsilon_a, \ldots, \epsilon_{a+b-1} \) only. ■

**Remark 3.1** From equations (22), (25), and (28) it follows that basing asymptotic inference for \( \rho \) on the limiting distribution of the studentized statistic requires knowledge of the underlying parameters \( \rho \) and \( \mu \) which defeats the purpose. Moreover, in the case \( |\rho| < 1 \) we also need to consistently estimate the dependence structure of the \( \{\epsilon_t\} \) process. This would be very difficult in practice as can be seen from equations (19) and (20). On the other hand, we have shown that subsampling works in all three cases and therefore allows to make asymptotically correct inference no matter what the underlying parameters and dependency structure of the \( \{\epsilon_t\} \) process are.

**Remark 3.2** If we limited ourselves to the case of \( |\rho| < 1 \) and i.i.d. innovations, asymptotic inference for \( \rho \) could also be based on the standard normality of the OLS \( t \)-statistic. However, it is known that finite sample properties of this approach are not good when \( \rho \) is close to one. The intuition here is that for \( \rho \) close to one the distribution of the \( t \)-statistic in a finite sample should not be all the different from its distribution for \( \rho = 1 \), a case which the normal method cannot handle. The subsampling approach, on the other hand, also works for \( \rho = 1 \) and we can therefore expect better finite sample properties. See Section 5 for some simulation studies that confirm this conjecture.
Remark 3.3 Note that unlike the bootstrap the subsampling method can handle discontinuities of the limiting distribution of (standardized or studentized) estimators as a function of the underlying model parameters. The intuition here is that the subsampling approximation of the sampling distribution of an estimator is based on subsample statistics computed from smaller blocks of the observed data. The subsample statistics are therefore always generated from the true model. The bootstrap, on the other hand, bases its approximation on pseudo statistics computed from "fake" data according to a bootstrap distribution estimated from the observed time series. The "fake" data will come from a model close to the truth, but not exactly the truth. This can cause the bootstrap to fail. A case in point is inference for the AR(1) parameter. The bootstrap works if $\rho < 1$, but it does not work if $\rho = 1$ (Basawa et al., 1991).

Remark 3.4 The assumption of strict stationarity of the innovation process $\{\epsilon_t\}$ can be relaxed to the extent that condition (ii) of Assumption 2.2 is still satisfied for the sampling distributions of the subsample statistics $\tau_b(\hat{\rho}_{b,a} - \rho)/\hat{\sigma}_{b,a}$. For instance, this would allow for seasonal effects reflected in nonconstant variances and/or changing distributions of the $\epsilon_t$.

Remark 3.5 In case one is willing to assume i.i.d. rather than uncorrelated innovations, the assumption of a finite $2 + \delta$ moment of $\epsilon_t$ in Proposition 3.1 is sufficient. Note also that in this case the mixing condition on the $\{X_t\}$ process is implied if the distribution of $\epsilon_t$ is continuous; see Doukhan (1995).

4 Choice of the Block Size

A practical issue in using the subsampling method is the choice of the block size $b$. Politis, Romano, and Wolf (1997) propose a calibration technique that corrects for over- or under-coverage of subsampling intervals for finite samples by adjusting the nominal confidence level accordingly. This technique involves generating pseudo sequences $X^*_1, \ldots, X^*_n$ using a suitable bootstrap method — Efron's (1979) bootstrap for i.i.d. data or Künsch's (1989) moving blocks bootstrap for time series data. Hence, this idea is limited to applications where standard bootstrap methods are also consistent and would fail for our problem at hand. We therefore present a different method which will work under more general conditions, that is, whenever subsampling applies.

Our method is of heuristic nature and we do not claim any optimality properties. It is based on the fact that for the subsampling method to be consistent the block size $b$ needs to tend to infinity with the sample size $n$ but a smaller rate, satisfying $b/n \to 0$. Indeed, for $b$ too close to $n$ all subsample statistics $\hat{\theta}_{n,i}$ will almost equal to $\hat{\theta}_n$, resulting in the subsampling distributions
$L_n$ or $K_n$ being too tight and in undercoverage of subsampling confidence intervals. If $b$ is too small, the intervals can undercut or overcover depending on the state of nature (e.g., Politis, Romano, and Wolf, 1997). This leaves a number of $b$ values in the “right range” where we would expect almost correct results, at least for big sample sizes. We exploit this idea by computing subsampling intervals for a large number of block sizes $b$ and then looking for a region where the intervals do not change very much. Within this region we then pick one interval according to some arbitrary criterion.

While this method can be carried out by “visual inspection” it is desirable to also have some automatic selection procedure, especially when simulation studies are to be carried out. The procedure we propose is based on minimizing a running standard deviation. Assume we compute subsampling intervals for block sizes $b$ in the range of $b_{\text{small}}$ to $b_{\text{big}}$. The endpoints of the confidence intervals should change in a smooth fashion as $b$ changes. A running standard deviation applied to the endpoints then determines the volatility around a specific $b$ value. We choose the value of $b$ with the smallest volatility. Here is a more formal description of the algorithm.

**Algorithm 4.1 (Minimizing Confidence Interval Volatility)**

1. For $b = b_{\text{small}}$ to $b = b_{\text{big}}$ compute a subsampling interval for $\theta$ at the desired confidence level, resulting in endpoints $I_{b,\text{low}}$ and $I_{b,\text{up}}$.

2. For each $b$ compute a volatility index $VI_b$ as the standard deviation of the interval endpoints in a neighborhood of $b$. More specifically, for a small integer $k$, let $VI_b$ be equal to the standard deviation of $\{I_{b-k,\text{low}}, \ldots, I_{b+k,\text{low}}\}$ plus the standard deviation of $\{I_{b-k,\text{up}}, \ldots, I_{b+k,\text{up}}\}$.

3. Pick the value $b^*$ with the smallest volatility index and report $[I_{b^*,\text{low}}, I_{b^*,\text{up}}]$ as the final confidence interval.

Some remarks concerning the implementation of this algorithm are in order.

**Remark 4.1** The range of $b$ values, determined by $b_{\text{small}}$ and $b_{\text{big}}$, which is included in the minimization algorithm is not very important, as long as it is not too narrow.

**Remark 4.2** To make the algorithm more computationally efficient, it might be desirable to skip a number of $b$ values in a regular fashion. For example, include only every other or every third $b$ between $b_{\text{small}}$ and $b_{\text{big}}$.

**Remark 4.3** The algorithm contains a model parameters $k$ Simulation studies have shown that the algorithm is very insensitive its choice. We usually employ $k = 2$ or $k = 3$. 


We now illustrate how the algorithm works with the help of two simulated data sets. First, we generated a time series of size \( n = 200 \) according to model (14), using \( \rho = 0.95, \mu = 0, \) and i.i.d. standard normal innovations. The range of \( b \) values was chosen as \( b_{\text{small}} = 10 \) and \( b_{\text{big}} = 36 \). The minimization of the volatility in step 2 was done using \( k = 2 \). The results are shown at the top of Figure 1. The left plot corresponds to equal-tailed confidence intervals while the right plot corresponds to symmetric confidence intervals. The block sizes \( b \) chosen by the algorithm are highlighted by a star. The resulting final confidence intervals are included in the plots together with the point estimate \( \hat{\rho}_n \).

This exercise was repeated for another data set of size \( n = 500 \), using \( \rho = 1, \mu = 0, \) and i.i.d. standard normal innovations. The range of \( b \) was there chosen as \( b_{\text{small}} = 15 \) and \( b_{\text{big}} = 58 \). The results are shown at the bottom of Figure 1.

From the plots it can be seen that symmetric intervals are somewhat more stable, that is, the endpoints change less as we vary \( b \). This behavior is typical and was observed for many other simulations as well.

5 Small Sample Performance

The purpose of this Section is to shed some light on the small sample performance of the subsampling method by means of a simulation study. As the data generating process we use our model (14). We employ \( \mu \) either equal to 0 or 0.1; the value 0.1 was chosen to make the time trend not overly dominating when \( \rho = 1 \). The AR(1) parameter \( \rho \) is one of the following: 1, 0.99, 0.95, 0.9, or 0.5. As innovation process \{\( \epsilon_t \)\} we consider i.i.d. standard normal, i.i.d. exponential with mean zero and variance one, and \( \epsilon_t = Z_t Z_{t-1} \) where the \( Z_t \) are i.i.d. standard normal. The last process is uncorrelated but one-dependent. As sample sizes we use \( n = 200 \) and \( n = 500 \).

Performance of confidence intervals is measured by empirical coverage probability of nominal two-sided 95\% intervals. We compute both equal-tailed and symmetric subsampling intervals, denoted by \( \text{Sub}_{ET} \) and \( \text{Sub}_{SYM} \), respectively. Estimated coverage is based on 1000 replications for each scenario.

We compare the subsampling method with the CLT and Stock's (1991) method. The CLT uses asymptotic standard normality of the \( t \)-statistic for \( \hat{\rho}_n \). This is the correct limiting distribution if \( |\rho| < 1 \) and the \( \epsilon_t \) are i.i.d. or if \( \rho = 1 \) and \( \mu \neq 1 \). The inference of Stock (1991) is limited to \( \rho \) values close to one, where "close" depends on the sample size. For \( n = 200 \) it works as low as \( \rho = 0.9 \) and for \( n = 500 \) it works as low as \( \rho = 0.95 \). Inference is based on the model 3 by rewriting it in the following way

\[
X_t = \hat{\alpha} + \hat{\delta}t + \rho X_{t-1} + \epsilon_t. \tag{30}
\]
The test statistic is \( \hat{\rho}^T \), the OLS t-statistic for \( \rho = 1 \) in (30). Part B of Table A.1 in Stock (1991) allows one to check, up to some mild interpolation, whether a certain \( \rho \) value is contained in the two-sided 95% confidence interval, given the outcome of \( \hat{\rho}^T \) and the sample size. For example, \( \rho = 1 \) is contained in the interval if \(-3.73 \leq \hat{\rho}^T \leq 0.67\), independent of the sample size. On the other hand, \( \rho = 0.99 \) is contained in the interval if \(-3.77 \leq \hat{\rho}^T \leq -0.85\) for \( n = 200 \) and if \(-3.87 \leq \hat{\rho}^T \leq -1.24\) for \( n = 500 \).

Our simulation results are presented in Tables 1 to 3 and show the following. As expected, the CLT does not work very well in cases where it should not, that is, if \( \rho = 1 \) or very close to one and \( \mu = 0 \), or if \( \rho < 1 \) and the innovations are dependent. However, it also does not work very well in the case \( \rho = 1 \) and \( \mu = 0.1 \) where it should. It does work well when \( \rho \) is not very close to one and the innovations are i.i.d.

The next thing to note is that equal-tailed subsampling intervals tend to undercover and are not competitive. On the other hand, symmetric subsampling intervals work rather well. Stock's intervals also work well when they apply, although they seem to be somewhat less robust against dependent innovations than subsampling. It also seems that they appear to consistently overcover somewhat for the case \( \rho = 1 \) which is in slight contrast to the simulation results reported in Stock (1991).

As a competitor to symmetric subsampling intervals that can be used universally one might think of the following method. Use Stock's intervals when they apply, that is, if \( \rho \) is close to one relative to the sample size and the CLT intervals otherwise. From the tables it can be seen that this is comparable to symmetric subsample intervals if the innovations are i.i.d.. However, it fails when the innovations are correlated, since the CLT intervals undercover significantly (which is not surprising, since they are known to asymptotically fail from our results in Section 3). Also, Stock's intervals are not quite as good as the subsampling intervals in this case.

6 Summary

In this paper we proposed a new method for constructing asymptotic confidence intervals for \( \rho \) in the basic AR(1) model

\[ X_t = \mu + \rho X_{t-1} + \epsilon_t, \]

where \( \{\epsilon_t\} \) is white noise and \( |\rho| \leq 1 \). This is a difficult problem, since the limiting distribution of the OLS estimator depends in a discontinuous way not only on \( \rho \) itself but also on \( \mu \) and the dependency structure of the innovations \( \epsilon_t \). For example, the discontinuity causes the standard bootstrap to fail.
The only method suggested so far (that we are aware of) is the local-to-unity asymptotics approach by Stock (1991). It assumes that $\rho$ shrinks to one as the sample size increases to infinity in the fashion $\rho = 1 + \frac{c}{n}$. The approach works well when $\rho$ is close to one (with respect to the sample size), but will fail for $\rho$ values moderately far away from one. This may be considered a "breakdown" problem.

Our method is based on the subsampling idea of Politis and Romano (1994). We approximate the sampling distribution of the OLS $t$-statistic for $\rho$ by recomputing it on smaller blocks of the observed data sequence. In those block $t$-statistics the unknown value of $\rho$ is replaced by $\hat{\rho}_n$, the estimator based on the entire sample. The empirical distribution of the block $t$-statistics serves as our approximation. Unlike bootstrapping, subsampling can handle the discontinuities of the limiting distribution of the $t$-statistic.

We provided an extension of previous theory needed for the case $\rho = 1$ when the $X_t$ sequence is no longer weakly dependent. Another extension was needed for studentized statistics computed from dependent data.

Some simulation studies showed that the finite sample properties of subsampling intervals are encouraging, at least when symmetric intervals are used. When $\rho$ is equal or close to one, their coverage is about as good as the coverage of Stock's intervals. When $\rho$ is far away from one, subsample intervals cover about as well as CLT intervals for i.i.d. innovations and much better than CLT intervals for uncorrelated but dependent innovations. The last observation is due to the often neglected fact that CLT intervals are inconsistent when innovations are dependent.

Finally, we would like to point out that subsampling is a very general and powerful technique and not restricted to models as simple as the AR(1) model discussed in this paper. For instance, if the goal was to find confidence intervals for the largest autoregressive root of a time series, we could base inference on the usual augmented Dicker-Fuller (1979) regression, which controls for short-term dynamics by including higher-order autoregressive terms in the regression. Subsampling will still apply, only the proofs would become more cumbersome due to the additional notation. Moreover, we are convinced that there are many other difficult econometric problems where subsampling could be used beneficially. Basically this is whenever the limiting distribution of an estimator depends on the underlying model parameters in a complicated and maybe even discontinuous way; one of many examples is inference in models with integrated or nearly integrated regressors as discussed in Elliot and Stock (1994), Cavanagh, Elliot, and Stock (1995), and Elliot (1998). The main condition for subsampling to work is that the underlying sequence or, which is more general, the subsample statistics are weakly dependent in a sufficient way.
References


Table 1: Estimated coverage probabilities of various nominal 95% subsampling confidence intervals when the innovations are i.i.d. standard normal. CLT denotes intervals which approximate the sampling distribution of the $t$-statistic by a standard normal distribution. Sub$_{ET}$ and Sub$_{SYM}$ denote equal-tailed and symmetric subsampling intervals, respectively. Stock denotes the intervals according to Stock (1991). The estimates are based on 1000 replications for each scenario.

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>CLT</th>
<th>Sub$_{ET}$</th>
<th>Sub$_{SYM}$</th>
<th>Stock</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.70</td>
<td>0.92</td>
<td>0.92</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td>0.99</td>
<td>0.82</td>
<td>0.91</td>
<td>0.93</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>0.95</td>
<td>0.89</td>
<td>0.85</td>
<td>0.94</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>0.90</td>
<td>0.93</td>
<td>0.84</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>0.50</td>
<td>0.95</td>
<td>0.87</td>
<td>0.93</td>
<td>NA</td>
<td>0.95</td>
</tr>
</tbody>
</table>

| $\mu = 0$, $n = 500$, $\epsilon_t = Z_t$ |
|-------|-----|------------|-------------|-------|--------|
| $\rho$ | CLT | Sub$_{ET}$ | Sub$_{SYM}$ | Stock | Target |
| 1     | 0.71| 0.95       | 0.95        | 0.98  | 0.95   |
| 0.99  | 0.87| 0.92       | 0.96        | 0.95  | 0.95   |
| 0.95  | 0.94| 0.84       | 0.96        | 0.96  | 0.95   |
| 0.90  | 0.95| 0.85       | 0.96        | NA    | 0.95   |
| 0.50  | 0.94| 0.90       | 0.94        | NA    | 0.95   |

| $\mu = 0.1$, $n = 200$, $\epsilon_t = Z_t$ |
|-------|-----|------------|-------------|-------|--------|
| $\rho$ | CLT | Sub$_{ET}$ | Sub$_{SYM}$ | Stock | Target |
| 1     | 0.79| 0.92       | 0.94        | 0.98  | 0.95   |
| 0.99  | 0.81| 0.91       | 0.93        | 0.95  | 0.95   |
| 0.95  | 0.88| 0.85       | 0.95        | 0.95  | 0.95   |
| 0.90  | 0.93| 0.84       | 0.95        | 0.95  | 0.95   |
| 0.50  | 0.95| 0.87       | 0.93        | NA    | 0.95   |

| $\mu = 0.1$, $n = 500$, $\epsilon_t = Z_t$ |
|-------|-----|------------|-------------|-------|--------|
| $\rho$ | CLT | Sub$_{ET}$ | Sub$_{SYM}$ | Stock | Target |
| 1     | 0.88| 0.91       | 0.97        | 0.98  | 0.95   |
| 0.99  | 0.87| 0.91       | 0.96        | 0.95  | 0.95   |
| 0.95  | 0.93| 0.86       | 0.97        | 0.95  | 0.95   |
| 0.90  | 0.94| 0.85       | 0.96        | NA    | 0.95   |
| 0.50  | 0.94| 0.90       | 0.94        | NA    | 0.95   |
Table 2: Estimated coverage probabilities of various nominal 95% subsampling confidence intervals when the innovations are i.i.d. exponential with mean zero and variance one. CLT denotes intervals which approximate the sampling distribution of the t-statistic by a standard normal distribution. Sub$_{ET}$ and Sub$_{SYM}$ denote equal-tailed and symmetric subsampling intervals, respectively. Stock denotes the intervals according to Stock (1991). The estimates are based on 1000 replications for each scenario.

<table>
<thead>
<tr>
<th>$\mu = 0$, $n = 200$, $\epsilon_t = E_t$</th>
<th>ClT</th>
<th>Sub$_{ET}$</th>
<th>Sub$_{SYM}$</th>
<th>Stock</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>0.68</td>
<td>0.91</td>
<td>0.92</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td>0.99</td>
<td>0.82</td>
<td>0.90</td>
<td>0.93</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>0.95</td>
<td>0.92</td>
<td>0.87</td>
<td>0.96</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>0.90</td>
<td>0.93</td>
<td>0.86</td>
<td>0.95</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td>0.50</td>
<td>0.95</td>
<td>0.83</td>
<td>0.92</td>
<td>NA</td>
<td>0.95</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\mu = 0$, $n = 500$, $\epsilon_t = E_t$</th>
<th>ClT</th>
<th>Sub$_{ET}$</th>
<th>Sub$_{SYM}$</th>
<th>Stock</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>0.69</td>
<td>0.95</td>
<td>0.93</td>
<td>0.98</td>
<td>0.95</td>
</tr>
<tr>
<td>0.99</td>
<td>0.88</td>
<td>0.90</td>
<td>0.96</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>0.95</td>
<td>0.93</td>
<td>0.84</td>
<td>0.97</td>
<td>0.96</td>
<td>0.95</td>
</tr>
<tr>
<td>0.90</td>
<td>0.95</td>
<td>0.83</td>
<td>0.96</td>
<td>NA</td>
<td>0.95</td>
</tr>
<tr>
<td>0.50</td>
<td>0.96</td>
<td>0.88</td>
<td>0.94</td>
<td>NA</td>
<td>0.95</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\mu = 0.1$, $n = 200$, $\epsilon_t = E_t$</th>
<th>ClT</th>
<th>Sub$_{ET}$</th>
<th>Sub$_{SYM}$</th>
<th>Stock</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>0.79</td>
<td>0.91</td>
<td>0.93</td>
<td>0.98</td>
<td>0.95</td>
</tr>
<tr>
<td>0.99</td>
<td>0.82</td>
<td>0.91</td>
<td>0.93</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>0.95</td>
<td>0.92</td>
<td>0.89</td>
<td>0.96</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>0.90</td>
<td>0.94</td>
<td>0.84</td>
<td>0.95</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td>0.50</td>
<td>0.95</td>
<td>0.83</td>
<td>0.91</td>
<td>NA</td>
<td>0.95</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\mu = 0.1$, $n = 500$, $\epsilon_t = E_t$</th>
<th>ClT</th>
<th>Sub$_{ET}$</th>
<th>Sub$_{SYM}$</th>
<th>Stock</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>0.86</td>
<td>0.90</td>
<td>0.97</td>
<td>0.98</td>
<td>0.95</td>
</tr>
<tr>
<td>0.99</td>
<td>0.89</td>
<td>0.91</td>
<td>0.97</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>0.95</td>
<td>0.94</td>
<td>0.83</td>
<td>0.96</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>0.90</td>
<td>0.95</td>
<td>0.84</td>
<td>0.96</td>
<td>NA</td>
<td>0.95</td>
</tr>
<tr>
<td>0.50</td>
<td>0.95</td>
<td>0.86</td>
<td>0.93</td>
<td>NA</td>
<td>0.95</td>
</tr>
</tbody>
</table>
Table 3: Estimated coverage probabilities of various nominal 95% subsampling confidence intervals when the innovations are $\epsilon_t = Z_t Z_{t-1}$, where the $Z_t$ are i.i.d. standard normal. CLT denotes intervals which approximate the sampling distribution of the $t$-statistic by a standard normal distribution. Sub$_{ET}$ and Sub$_{SYM}$ denote equal-tailed and symmetric subsampling intervals, respectively. Stock denotes the intervals according to Stock (1991). The estimates are based on 1000 replications for each scenario.

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>CLT</th>
<th>Sub$_{ET}$</th>
<th>Sub$_{SYM}$</th>
<th>Stock</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.70</td>
<td>0.96</td>
<td>0.95</td>
<td>0.98</td>
<td>0.95</td>
</tr>
<tr>
<td>0.99</td>
<td>0.82</td>
<td>0.96</td>
<td>0.96</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>0.95</td>
<td>0.88</td>
<td>0.88</td>
<td>0.96</td>
<td>0.93</td>
<td>0.95</td>
</tr>
<tr>
<td>0.90</td>
<td>0.88</td>
<td>0.84</td>
<td>0.94</td>
<td>0.92</td>
<td>0.95</td>
</tr>
<tr>
<td>0.50</td>
<td>0.81</td>
<td>0.77</td>
<td>0.90</td>
<td>NA</td>
<td>0.95</td>
</tr>
</tbody>
</table>

<p>| $\mu = 0, n = 500, \epsilon_t = Z_t Z_{t-1}$ |
|-------|------|------------|-------------|-------|--------|</p>
<table>
<thead>
<tr>
<th>$\rho$</th>
<th>CLT</th>
<th>Sub$_{ET}$</th>
<th>Sub$_{SYM}$</th>
<th>Stock</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.69</td>
<td>0.96</td>
<td>0.95</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td>0.99</td>
<td>0.87</td>
<td>0.92</td>
<td>0.97</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>0.95</td>
<td>0.91</td>
<td>0.84</td>
<td>0.96</td>
<td>0.93</td>
<td>0.95</td>
</tr>
<tr>
<td>0.90</td>
<td>0.92</td>
<td>0.85</td>
<td>0.97</td>
<td>NA</td>
<td>0.95</td>
</tr>
<tr>
<td>0.50</td>
<td>0.79</td>
<td>0.82</td>
<td>0.91</td>
<td>NA</td>
<td>0.95</td>
</tr>
</tbody>
</table>

<p>| $\mu = 0.1, n = 200, \epsilon_t = Z_t Z_{t-1}$ |
|-------|------|------------|-------------|-------|--------|</p>
<table>
<thead>
<tr>
<th>$\rho$</th>
<th>CLT</th>
<th>Sub$_{ET}$</th>
<th>Sub$_{SYM}$</th>
<th>Stock</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.80</td>
<td>0.94</td>
<td>0.96</td>
<td>0.96</td>
<td>0.95</td>
</tr>
<tr>
<td>0.99</td>
<td>0.81</td>
<td>0.95</td>
<td>0.96</td>
<td>0.91</td>
<td>0.95</td>
</tr>
<tr>
<td>0.95</td>
<td>0.90</td>
<td>0.88</td>
<td>0.97</td>
<td>0.91</td>
<td>0.95</td>
</tr>
<tr>
<td>0.90</td>
<td>0.87</td>
<td>0.84</td>
<td>0.95</td>
<td>0.93</td>
<td>0.95</td>
</tr>
<tr>
<td>0.50</td>
<td>0.82</td>
<td>0.79</td>
<td>0.91</td>
<td>NA</td>
<td>0.95</td>
</tr>
</tbody>
</table>

<p>| $\mu = 0.1, n = 500, \epsilon_t = Z_t Z_{t-1}$ |
|-------|------|------------|-------------|-------|--------|</p>
<table>
<thead>
<tr>
<th>$\rho$</th>
<th>CLT</th>
<th>Sub$_{ET}$</th>
<th>Sub$_{SYM}$</th>
<th>Stock</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.86</td>
<td>0.93</td>
<td>0.97</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td>0.99</td>
<td>0.88</td>
<td>0.93</td>
<td>0.97</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>0.95</td>
<td>0.91</td>
<td>0.85</td>
<td>0.97</td>
<td>0.93</td>
<td>0.95</td>
</tr>
<tr>
<td>0.90</td>
<td>0.90</td>
<td>0.82</td>
<td>0.96</td>
<td>NA</td>
<td>0.95</td>
</tr>
<tr>
<td>0.50</td>
<td>0.78</td>
<td>0.82</td>
<td>0.91</td>
<td>NA</td>
<td>0.95</td>
</tr>
</tbody>
</table>
Figure 1: Illustration of the Minimizing Confidence Interval Volatility Algorithm for two data sets. The plots on the left correspond to equal-tailed confidence intervals, while the plots on the right correspond to symmetric confidence intervals. The block sizes selected by the algorithm are highlighted by a star. The final confidence intervals appear within the plots together with the point estimates.