Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 138 (2008) 3858–3868
journal of statistical planning and inference
www.elsevier.com/locate/jspi

A small sample confidence interval for autoregressive parameters

Colin Gallagher*, Ferebee Tunno

Department of Mathematical Sciences, Clemson University, Box 340975, Clemson, SC 29634-0975, USA

Received 16 February 2007; received in revised form 17 January 2008; accepted 4 February 2008
Available online 14 February 2008

Abstract

This paper is concerned with interval estimation of an autoregressive parameter when the parameter space allows for magnitudes outside the unit interval. In this case, intervals based on the least-squares estimator tend to require a high level of numerical computation and can be unreliable for small sample sizes. Intervals based on the asymptotic distribution of instrumental variable estimators provide an alternative. If the instrument is taken to be the sign function, the interval is centered at the Cauchy estimator and a large sample interval can be created by estimating the standard error of this estimator. The interval proposed in this paper avoids estimating this standard error and results in a small sample improvement in coverage probability. In fact, small sample coverage is exact when the innovations come from a normal distribution.

Keywords: Unit root; Linear trend; Small sample; Instrumental variable

1. Introduction

Consider a time series with trend defined by

$$Y_t = \mu(t) + X_t, \tag{1.1}$$

where the X_t's may either follow the stationary AR(1) model

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \pm 3, \ldots, \quad |\phi| < 1 \tag{1.2}$$

with observations $X_1, \ldots, X_n$ or the non-stationary AR(1) model

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad X_0 = 0, \quad t = 1, 2, \ldots, n, \quad \phi \in \mathbf{R}. \tag{1.3}$$

We assume that both the stationary and non-stationary AR(1) models have independent identically distributed (iid) errors possessing a density symmetric about zero, with $E(\varepsilon_t^q) < \infty$ for some $q > 2$.

This paper focuses on interval estimation of the autoregressive parameter ϕ. More specifically, our goal is to investigate simple intervals which require no numerical approximation. Applied statisticians might be tempted to use the

* Corresponding author. Tel.: +1 864 656 2286.
E-mail address: cgallag@clemson.edu (C. Gallagher).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2008.02.001

least-squares estimator (OLSE)

$$\widetilde{\phi}_1 = \frac{\sum_{t=2}^{n} X_{t-1}X_t}{\sum_{t=2}^{n} X_{t-1}^2} \tag{1.4}$$

to construct confidence intervals for ϕ centered at $\widetilde{\phi}_1$ because of the asymptotic distribution of this estimator: for $|\phi| < 1$, $\sqrt{n}(\widetilde{\phi}_1 - \phi)$ is asymptotically normal (Mann and Wald, 1943); for $|\phi| > 1$, $\phi^n(\widetilde{\phi}_1 - \phi)/(\phi^2 - 1)$ has asymptotic Cauchy Distribution provided the error distribution is normal (White, 1958); when $|\phi| = 1$, $n(\widetilde{\phi}_1 - \phi)$ has limiting distribution which can be expressed as the ratio of two functionals of standard Brownian Motion (Rao, 1978).

There are several reasons, however, why basing confidence intervals on the asymptotic distribution of $\widetilde{\phi}_1$ is not necessarily a good idea. It has long been known that the least-squares estimators of autoregressive coefficients can have large bias; see, for example, Marriott and Pope (1954), Shaman and Stine (1988), and Newbold and Agiakloglou (1993). Also, making a confidence interval when $\widetilde{\phi}_1$ is near unity usually requires considering parameter values crossing the three limiting cases given above, which is especially problematic in the second case as the limiting distribution depends on the distribution of the errors. Furthermore, even in the stationary case, the asymptotic normal distribution provides a poor approximation for small to moderate sample sizes when $|\phi|$ is near 1 (Ahtola and Tiao, 1984). Most attempts at improving interval estimation for ϕ are based on somehow modifying the asymptotically optimal least-squares estimator or by attempting to better approximate its finite sample distribution.

A variety of methods have been proposed to improve the finite sample distributional approximation when the errors follow a normal distribution. Phillips (1978) derived a saddlepoint approximation to the probability density function which is very accurate in the center of the distribution and Wang (1992) and Lieberman (1994) derive accurate saddle point approximations to the distribution function. In the local to unity case, Perron (1991) provides a very accurate approximation which is robust against fixed initial values x_0. More recently, Ali (2002) gives a nice review of these approximations and proposes using a uniform asymptotic expansion of the distribution of $\widetilde{\phi}_1$ in terms of the joint characteristic function of the numerator and denominator of the estimator. Ali gives evidence that his saddlepoint approximation method works well in the asymptotic stationary, unit root and explosive root cases. These sophisticated methods tend to have two major drawbacks however. First, they are based on assuming normality and are not necessarily robust to departures from normality. Second, they require a good deal of computational approximation, and thus may not be attractive to the applied practitioner.

Of course, other computational methods can be used to create accurate confidence intervals. To improve small sample performance one can typically use bootstrapping. Basawa et al. (1991) consider standard bootstrap intervals for $|\phi| = 1$ and show that in this case the bootstrap is asymptotically invalid. This result remains true in the local to unity case, e.g., $\phi_n = 1 - c/n$, indicating that the naïve bootstrap performs poorly when $|\phi|$ is near 1. Bootstrap intervals can be improved, for example, by using the test-inversion bootstrap intervals of Carpenter (1999) or using the grid bootstrap of Hansen (1999). Elliot and Stock (2001) take a different approach. They consider inverting a sequence of asymptotically point optimal tests and create intervals by numerical approximation of corresponding characteristic functions. Again, although these methods can improve accuracy, they also require a fair bit of computational work.

Another tack was taken by Fuller (1996) and Andrews (1993) who found a way to circumvent the bias issues associated with $|\phi| \approx 1$ altogether by constructing confidence intervals based around median-unbiased estimators. Andrews considered these estimators, which can be used to find finite sample intervals under the normal model; ϕ is estimated by numerically inverting a median function which depends on ϕ, and quantiles corresponding to 90% confidence intervals are given for various n and ϕ values.

So and Shin (1999) used what they call a "Cauchy estimator"

$$\widetilde{\phi}_0 = \frac{\sum_{t=2}^{n} S_{t-1}X_t}{\sum_{t=2}^{n} |X_{t-1}|} \tag{1.5}$$

to construct intervals that have a shorter average length than those of Andrews, where S_t is the sign function of X_t. So and Shin note that $\widetilde{\phi}_0$ is approximately median unbiased as well. They obtained their pivotal statistic for the interval's construction by showing that

$$Z_0 = \frac{(\widetilde{\phi}_0 - \phi)}{\mathrm{se}(\widetilde{\phi}_0)} = \frac{1}{\hat{\sigma}\sqrt{n}} \sum_{t=2}^{n} S_{t-1}\varepsilon_t$$

is asymptotically normal, where $\hat{\sigma}$ is a consistent estimator of σ. Phillips et al. (2004) showed that the Cauchy estimator has asymptotically optimal precision properties in a certain class of instrumental variable estimators. Both of these papers conclude that intervals based on the Cauchy estimator have good large sample properties.

In this paper, we will show that for symmetric errors the distribution of $\sum_{t=2}^{n} S_{t-1}\varepsilon_t$ is the same as the distribution of the partial sums $\sum_{t=2}^{n}\varepsilon_t$ for finite n and take advantage of this fact to construct confidence intervals for ϕ, extending our work to time series models possessing a parametric trend. The simple interval given in this paper is centered at a weighted average of $\widetilde{\phi}_0$ and $\widetilde{\phi}_1$, and requires no numerical computation or approximation. For finite sample sizes n, it yields exact coverage probabilities under models (1.2) and (1.3) when the error terms follow a normal distribution. Since the proposed interval is based on self-normalized sums of the error terms, the coverage properties remain very good for non-normal, but symmetric errors; self-normalized sums of iid symmetric random variables converge very quickly to a standard normal distribution (Efron, 1969; Bentkus and Götze, 1996). The end points of the interval we develop below converge very quickly to the endpoints of the interval given in So and Shin (1999) and the interval given by formula (27) in Phillips et al. (2004); see Fig. 4. For small sample sizes the proposed interval has slightly better coverage probability. Thus the interval given in this paper can be seen as a small sample correction of the intervals investigated in So and Shin (1999) and Phillips et al. (2004).

Since the proposed interval will inherit the asymptotic properties of So and Shin's interval, we focus our attention on the finite sample properties of the interval. We begin by putting both $\widetilde{\phi}_0$ and the least-squares estimator $\widetilde{\phi}_1$ in a wider context by revealing them to be special cases of a more general weighted least-squares estimator.

2. Weighted least-squares estimators

Consider the sign function defined by

$$S_t = \begin{cases} 1, & X_t > 0, \\ 0, & X_t = 0, \\ -1, & X_t < 0. \end{cases}$$

If we let

$$X_t^{\langle p \rangle} = S_t |X_t|^p$$

then, given observations $X_1, X_2, \ldots, X_n$,

$$\widetilde{\phi}_p = \frac{\sum_{t=2}^{n} X_{t-1}^{\langle p \rangle} X_t}{\sum_{t=2}^{n} |X_{t-1}|^{p+1}}$$

can be thought of as a weighted least-squares estimator with form

$$\frac{\sum_{t=2}^{n} W_t X_t X_{t-1}}{\sum_{t=2}^{n} W_t X_{t-1}^2} \quad \text{and} \quad \text{weight } W_t = |X_{t-1}|^{p-1}.$$

Notice that when $p=1$, we obtain the ordinary (unweighted) least-squares estimator and when $p=0$, we get the Cauchy estimator. $\widetilde{\phi}_p$ can also be thought of as an instrumental variable estimator, as defined by Phillips et al. (2004). Here the instrument generating function $F(x) = x^{\langle p \rangle}$ has asymptotic order λ^p and limit homogeneous function $H(x) = x^{\langle p \rangle}$.

The following result can be proved using martingale difference central limit theory or as a corollary of Theorem 5.1 from Phillips et al. (2004).

Theorem 2.1. *Let* $r = \max(2p, p+1)$. *Under both models* (1.2) *and* (1.3), *for all p such that* $\lim_{t\to\infty} E(X_t^r) < \infty$, *we have*

$$\sqrt{n}(\widetilde{\phi}_p - \phi) \xrightarrow{\mathrm{D}} \mathrm{N}\left(0, \operatorname*{plim}_{n\to\infty} \frac{\sigma^2 n^{-1}\sum_{t=2}^{n} X_{t-1}^{2p}}{(n^{-1}\sum_{t=2}^{n} |X_{t-1}|^{p+1})^2}\right)$$

when $|\phi| < 1$ (*see Appendix for proof*).

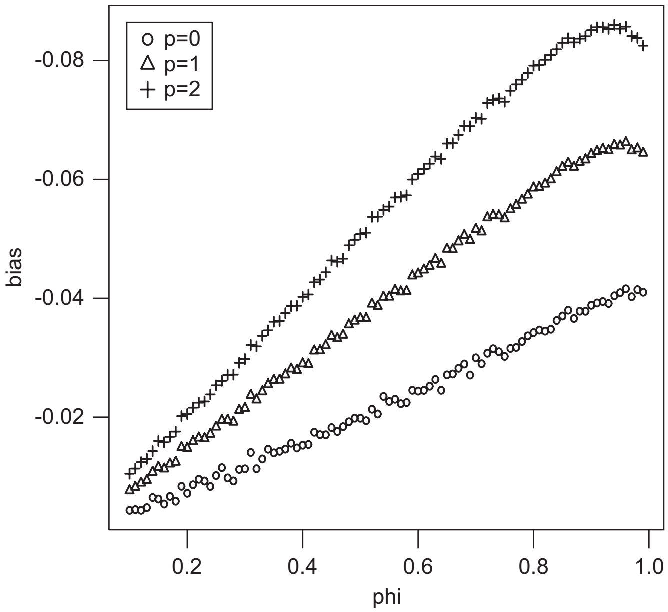

Fig. 1. Simulated bias.

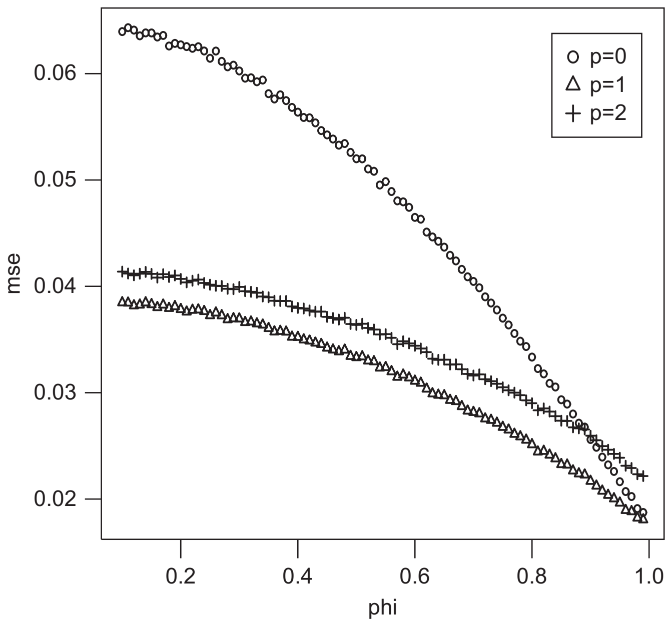

Fig. 2. Simulated MSE.

Clearly $\widetilde{\phi}_p$ is a consistent and asymptotically normal estimator of ϕ, regardless of the value of p. Finite sample properties of $\widetilde{\phi}_p$, however, are indeed a function of p.

Figs. 1 and 2 show bias and mean squared error (MSE) results of a simulation study under model (1.3) with $0 < \phi \leqslant 1$ and $\varepsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$. Each point comes from the average of 100,000 estimates, with $n = 25$ per estimate. In Fig. 1, bias is measured as a function of ϕ, for $p = 0, 1, 2$. As ϕ approaches zero, the bias for $p = 0, 1, 2$ does as well, but as ϕ approaches 1, the values fan out. Notice that the Cauchy estimator $\widetilde{\phi}_0$ has the smallest absolute bias.

Fig. 2 shows MSE as a function of $0 < \phi \leqslant 1$, for $p = 0, 1, 2$. As ϕ approaches 1, the MSE approaches zero for $p = 0, 1, 2$. The Cauchy estimator $\widetilde{\phi}_0$ starts off with a disproportionately larger MSE, but closes the gap as ϕ heads towards 1. The MSE of both $\widetilde{\phi}_2$ and the OLSE $\widetilde{\phi}_1$ decrease at the same rate, with $\widetilde{\phi}_1$ having smaller MSE. Based on these simulation results, and many not included in this paper, it appears that the Cauchy estimator minimizes finite sample bias, while the ordinary least-squares estimator seems to have smaller finite sample MSE.

3. Interval construction

3.1. AR(1) model

The result of the following theorem, combined with the desirable properties of both $\widetilde{\phi}_0$ and $\widetilde{\phi}_1$, can be used to create confidence intervals for ϕ when we have a simple AR(1) process such as those found in models (1.2) and (1.3). In fact, the pivotal statistic will end up having an *exact* distribution for finite n when $\{\varepsilon_t\} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$, and will converge very quickly to a standard normal when the errors are symmetric about zero.

Theorem 3.1. *Under both models* (1.2) *and* (1.3), *we have*

$$\{\varepsilon_2 S_1, \varepsilon_3 S_2, \ldots, \varepsilon_n S_{n-1}\} \sim \{\varepsilon_2, \varepsilon_3, \ldots, \varepsilon_n\}$$

for all $n \geqslant 2$ (*see Appendix for proof*).

Given observations $X_1, X_2, \ldots, X_n$, let $\Gamma_t = \varepsilon_t S_{t-1}$ and $\overline{\Gamma} = \sum_{t=2}^{n} \Gamma_t/(n-1)$. Then, if we let

$$T = \frac{\sqrt{n-2}\,\overline{\Gamma}}{\sqrt{\frac{1}{n-1}\sum_{t=2}^{n} \Gamma_t^2 - \overline{\Gamma}^2}}$$

and

$$R = \frac{\sum_{t=2}^{n} \Gamma_t}{\sqrt{\sum_{t=2}^{n} \Gamma_t^2}}, \tag{3.1}$$

it can easily be shown that

$$T = \frac{R\sqrt{n-2}}{\sqrt{n-1-R^2}}$$

(Efron, 1969). If we assume that $\{\varepsilon_t\} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$, then $\Gamma_2, \Gamma_3, \ldots, \Gamma_n \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$, which implies that T follows a Student's distribution with $n-2$ degrees of freedom (i.e., $T \sim t_{n-2}$).

If we wish to create a $(1-\alpha) \times 100\%$ confidence interval for ϕ, then observe that

$$\begin{aligned} 1-\alpha &= P(|T| < t_{n-2,\alpha/2}) = P(|R| < x) \\ &= P\left(\frac{(\sum_{t=2}^{n} \varepsilon_t S_{t-1})^2}{\sum_{t=2}^{n} \varepsilon_t^2} < x^2\right), \end{aligned}$$

where

$$x = \frac{t_{n-2,\alpha/2}\sqrt{n-1}}{\sqrt{n-2+t_{n-2,\alpha/2}^2}}$$

and $t_{n-2,\alpha/2}$ is the percentile with an area of $\alpha/2$ to its right. Noticing that $\varepsilon_t S_{t-1} = X_t S_{t-1} - \phi|X_{t-1}|$, and after some algebra we find that the above probability can be expressed in terms of a quadratic function in ϕ:

$$1-\alpha = P(a\phi^2 + b\phi + c < 0),$$

where

$$c=\widetilde{\phi}_0^2-x^2\sum_{t=2}^{n}X_t^2\bigg/\left(\sum_{t=2}^{n}|X_{t-1}|\right)^2,\quad b=2(x^2\widetilde{\phi}_1 r-\widetilde{\phi}_0),$$

$$a=1-x^2r\quad\text{and}\quad r=\frac{\sum_{t=2}^{n}X_{t-1}^2}{(\sum_{t=2}^{n}|X_{t-1}|)^2}. \tag{3.2}$$

Provided that $a>0$ and $b^2-4ac>0$ (see Appendix), a $(1-\alpha)\times 100\%$ confidence interval for ϕ has endpoints

$$\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

which are functions of both $\widetilde{\phi}_0$ and $\widetilde{\phi}_1$. In particular, the center of the interval is a weighted average of $\widetilde{\phi}_0$ and $\widetilde{\phi}_1$:

$$-\frac{b}{2a}=\left(\frac{1}{1-rx^2}\right)\widetilde{\phi}_0+\left(\frac{-rx^2}{1-rx^2}\right)\widetilde{\phi}_1.$$

3.2. AR(1) model with parametric trend

For ease of presentation, let us consider an AR(1) model with linear trend:

$$Y_t=\beta_0+\beta_1 t+X_t. \tag{3.3}$$

Under this model, the detrended observations, $\hat{X}_t=Y_t-\hat{Y}_t$, can be used to create a confidence interval for ϕ. For general linear models one can use *sequential* least-squares and obtain $\hat{Y}_t$ by regressing Y_j on j for $j\leqslant t$. This will keep $\widehat{X}_t$ and ε_{t+1} independent. Specifically, for a linear trend we can use the recursive detrending scheme given in Phillips et al. (2004). In this case one can simply replace X_t and X_{t-1} in (1.4), (1.5), and (3.2) with,

$$\hat{X}_t=Y_t-\frac{1}{n-t+1}(Y_n-Y_{t-1})+\frac{2}{t-1}\sum_{i=1}^{t-1}Y_i-\frac{6}{t(t-1)}\sum_{i=1}^{t-1}iY_i, \tag{3.4}$$

$$\hat{X}_{t-1}=Y_{t-1}+\frac{2}{t-1}\sum_{i=1}^{t-1}Y_i-\frac{6}{t(t-1)}\sum_{i=1}^{t-1}iY_i \tag{3.5}$$

for $t\geqslant 2$. We have

$$\varepsilon_t S_{t-1}\approx(\widehat{X}_t-\phi\widehat{X}_{t-1})\widehat{S}_{t-1}=\widehat{X}_t\widehat{S}_{t-1}-\phi|\widehat{X}_{t-1}|.$$

This approximate relationship can be exploited just as before to obtain confidence intervals for ϕ.

4. Simulations

In this section, we investigate the small sample properties of the interval developed in this paper and compare it to the intervals from So and Shin (1999) and a version of the interval given in Phillips et al. (2004), both of which are centered at the Cauchy estimator and have form

$$\tilde{\phi}_0\pm s(\tilde{\phi}_0)z_{\alpha/2},$$

where $z_{\alpha/2}$ is the standard normal quantile with an area of $\alpha/2$ to its right. Following So and Shin (1999) we took

$$s(\tilde{\phi}_0)=\frac{\sqrt{n}}{\sum_{t=2}^{n}|X_{t-1}|}\sqrt{\sum_{t=2}^{n}(X_t-\widetilde{\phi}_1X_{t-1})^2/(n-2)}.$$

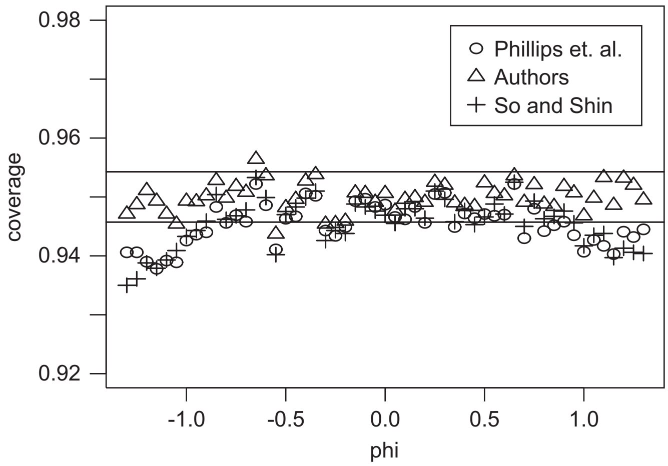

Fig. 3. Coverage probabilities, when $n = 16$, for model (1.3) with 10,000 simulations for each $\phi \in (-1.25, 1.25)$. The horizontal band is comprised of pointwise bounds used in testing the null hypothesis that the true coverage probability of the interval is 0.95, with type I error $\alpha = 0.05$.

The formulas given in Phillips et al. (2004) indicate one should take

$$s(\tilde{\phi}_0) = \hat{\sigma}\frac{\sqrt{\sum_{t=2}^{n} S_{t-1}^2}}{\sum_{t=2}^{n}|X_{t-1}|},$$

where $\hat{\sigma}^2$ is any consistent estimator of σ^2. We took $\hat{\sigma}^2$ to be $\sum_{t=1}^{n}(X_t - \widetilde{\phi}_0 X_{t-1})^2/(n-1)$. In all simulations discussed here, we assume $\{\varepsilon_t\} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$ and consider 95% intervals. For symmetric non-normal errors, the results are similar. In all simulations, we took $\phi \in (-1 - 1/\sqrt{n}, 1 + 1/\sqrt{n})$ and simulated under models (1.3), and (3.3) with (1.3). For each ϕ, we simulated 10,000 intervals using each method and calculated empirical coverage probabilities and average lengths for all intervals considered. Under the model with linear trend, we considered detrending using (3.4) as well as sequential least-squares.

Based on many simulations we conclude that under the simple AR(1) models and small n (say $n < 100$), the proposed interval is slightly wider than the other two intervals, and thus has simulated coverage closer to the nominal level. The relative coverage is exemplified in Fig. 3 which graphs simulated coverage as a function of ϕ under model (1.3) when $n = 16$. We see that for several ϕ values the large sample intervals have simulated coverage statistically smaller than the nominal 95% level. Also, the proposed small sample interval has higher simulated coverage probability than either of the large sample intervals. Not surprisingly, this larger coverage probability comes at the cost of interval length. The average length of the simulated intervals for $n = 16$ are plotted in part (a) of Fig. 4. The Phillips et al. interval tends to be the narrowest, the interval based on So and Shin tends to be slightly wider and the proposed small sample interval is of course the widest. As the sample size increases, the three intervals become indistinguishable. This can be seen in Fig. 4 part (b), which indicates that by the time $n = 100$, the intervals are essentially the same. For $16 < n < 100$, simulations not included here indicate that the coverage improvement becomes less pronounced as n increases. For the simple AR(1) models, the increase in coverage probability is less pronounced than under the model with linear trend.

Small sample time series trend estimates can be unreliable in the presence of autocorrelation, and so we expect some loss of coverage due to difficulty in detrending the data. Generally, the improvement of the proposed small sample interval is more pronounced under linear trend. We show results of simulations under model (3.3) for $n = 25$ in Fig. 5. The points show results for intervals created using the detrending scheme given in (3.4) for $\phi \in (-1.05, 1.05)$. The pattern is a result of this detrending scheme, which behaves well for ϕ in the vicinity of 1, but apparently behaves strangely for ϕ near 0.6. The detrending (3.4) also worked poorly for $|\phi| > 1.1$ with the proposed interval having the highest coverage, but well below the nominal 95% level. Notice also that in Fig. 5 the intervals tend to be too wide (overcover) for $\phi \leqslant -1$.

We also consider detrending using sequential least-squares estimates for the trend parameters. For both So and Shin (1999) and Phillips et al. (2004) intervals, detrending via sequential least-squares resulted in lower coverage probability

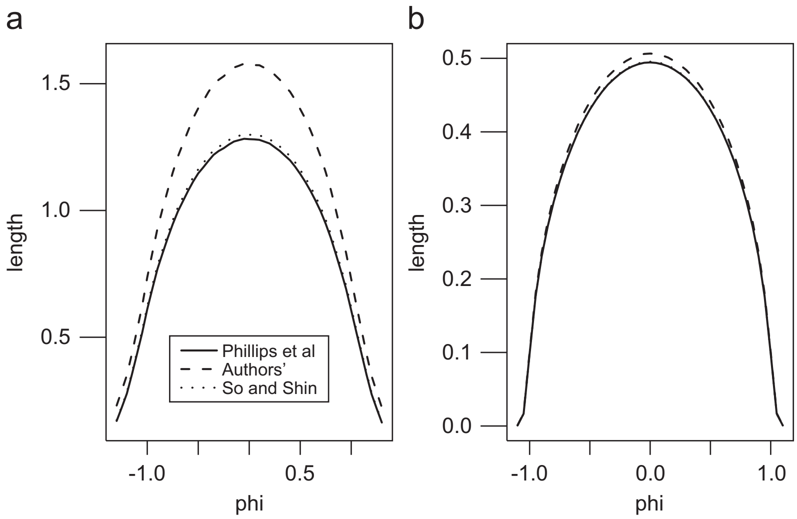

Fig. 4. Average interval lengths for model (1.3) when (a) $n = 16$ and (b) $n = 100$.

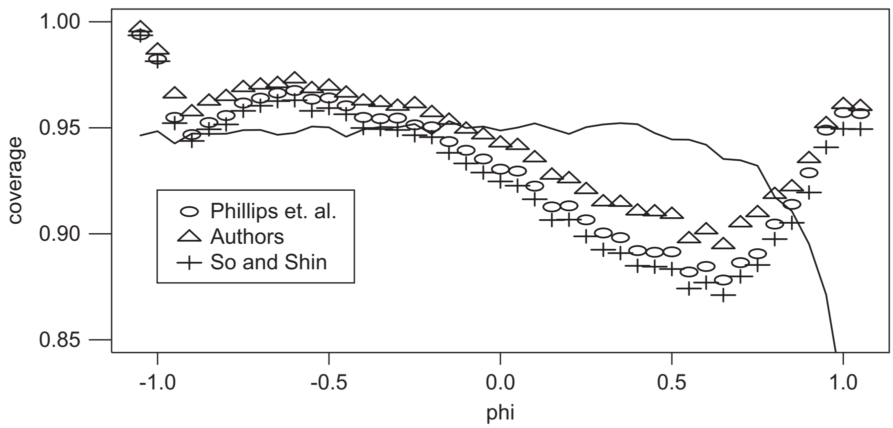

Fig. 5. Simulated interval coverage using non-stationary AR(1) models with linear trend and $n = 25$.

than detrending using (3.4), so these results are not included. However, the curve in Fig. 5 indicates simulated coverage of the proposed interval using sequential least-squares to detrend the data; using this scheme, the coverage is quite good for $\phi \leqslant 0.5$. The poor coverage as ϕ approaches 1 is due to the effect of correlation on least-squares estimates of time series trends.

5. Conclusions

In this paper, we have developed a small sample interval for the autoregressive parameter. Like the interval proposed by So and Shin, the interval works well for $|\phi| < 1$ as well as for $|\phi| \geqslant 1$. Under the assumption of normality, the interval has exact (as opposed to asymptotic) coverage. For small sample sizes, the proposed interval has higher coverage probabilities than either of the two intervals based on the asymptotic distribution of the Cauchy estimator, both of which tend to have smaller than nominal coverage. This higher coverage is of course at the cost of interval length. As the sample size grows, the endpoints of the proposed small sample interval tend very quickly to those based on So and Shin (1999). In fact, careful inspection of the authors' interval indicates that it should be asymptotically equivalent to that of So and Shin. The self-normalized sum (3.1) and the So and Shin pivot converge to the same random variable. Thus, the interval proposed can be seen as a finite sample correction for intervals based on the asymptotic distribution of the Cauchy estimator, such as those from So and Shin (1999) and Phillips et al. (2004).

Acknowledgments

The authors wish to express their gratitude to the two anonymous referees who took the time to carefully read our paper and provide us with valuable comments.

Appendix A. Proof of Theorem 2.1

Consider model (1.2) and observe that

$$\sqrt{n}(\widetilde{\phi}_p - \phi)\frac{\sum_{t=2}^n |X_{t-1}|^{p+1}}{n} = \frac{\sum_{t=2}^n \varepsilon_t X_{t-1}^{\langle p\rangle}}{\sqrt{n}}.$$

Let $\mathscr{F}_t$ be the σ-field generated by $\{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$, where $\varepsilon_t X_{t-1}^{\langle p\rangle} \in \mathscr{F}_t$ and

$$E(\varepsilon_t X_{t-1}^{\langle p\rangle}|\mathscr{F}_{t-1}) = 0.$$

Then $\{\varepsilon_t X_{t-1}^{\langle p\rangle}\}$ is a martingale difference sequence with respect to $\{\mathscr{F}_t\}$. If we define

$$V_n = \sum_{t=2}^n E[(\varepsilon_t X_{t-1}^{\langle p\rangle})^2|\mathscr{F}_{t-1}] = \sum_{t=2}^n X_{t-1}^{2p}\sigma^2$$

then

$$\begin{aligned}\frac{V_n}{n} &= \sigma^2\frac{\sum_{t=2}^n X_{t-1}^{2p}}{n} \xrightarrow{\mathrm{P}} \sigma^2 E(X_{t-1}^{2p})\\ &\Rightarrow \frac{\sum_{t=2}^n \varepsilon_t X_{t-1}^{\langle p\rangle}}{\sqrt{n}} \xrightarrow{\mathrm{D}} \mathrm{N}(0, \sigma^2 E(X_{t-1}^{2p}))\end{aligned}$$

by the martingale central limit theorem (see Durrett, 1996, p. 417). Thus, because

$$\frac{\sum_{t=2}^n |X_{t-1}|^{p+1}}{n} \xrightarrow{\mathrm{P}} E|X_{t-1}|^{p+1}$$

we then have

$$\sqrt{n}(\widetilde{\phi}_p - \phi) \xrightarrow{\mathrm{D}} \mathrm{N}\left(0, \frac{\sigma^2 E(X_{t-1}^{2p})}{(E|X_{t-1}|^{p+1})^2}\right).$$

The same result holds for model (1.3) by a similar argument.

Appendix B. Proof of Theorem 3.1

The result follows easily from the independence and symmetry of the sequence $\{\varepsilon_t\}$. We give a direct proof using the principal of mathematical induction. First note $P(\varepsilon_2 S_1 \leqslant c_2) = P(\varepsilon_2 \leqslant c_2)$. Now, assume that

$$\varepsilon_2 S_1, \varepsilon_3 S_2, \ldots, \varepsilon_{k-1} S_{k-2} \sim \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_{k-1}$$

for some positive integer k. Using symmetry and the fact that ε_k is independent of $X_1, X_2, \ldots, X_{k-1}$, we have

$$\begin{aligned} P(\varepsilon_2 S_1 \leqslant c_2, \ldots, \varepsilon_k S_{k-1} \leqslant c_k) &= P(\varepsilon_2 S_1 \leqslant c_2, \ldots, \varepsilon_k S_{k-1} \leqslant c_k, S_{k-1} = -1) \\ &\quad + P(\varepsilon_2 S_1 \leqslant c_2, \ldots, \varepsilon_k S_{k-1} \leqslant c_k, S_{k-1} = 1) \\ &= P(\varepsilon_2 S_1 \leqslant c_2, \ldots, \varepsilon_{k-1} S_{k-2} \leqslant c_{k-1}, S_{k-1} = 1) P(\varepsilon_k \leqslant c_k) \\ &\quad + P(\varepsilon_2 S_1 \leqslant c_2, \ldots, \varepsilon_{k-1} S_{k-2} \leqslant c_{k-1}, S_{k-1} = -1) P(-\varepsilon_k \leqslant c_k) \\ &= P(\varepsilon_2 \leqslant c_2, \ldots, \varepsilon_{k-1} \leqslant c_{k-1}, \varepsilon_k \leqslant c_k). \end{aligned}$$

Thus, by induction, we have

$$\varepsilon_2 S_1, \varepsilon_3 S_2, \ldots, \varepsilon_n S_{n-1} \sim \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_n \quad \text{for all } n \geqslant 2.$$

Appendix C. Discussion of necessary quadratic conditions

Let $V_1, V_2, \ldots, V_{n-1}$ be the order statistics for $|X_1|, |X_2|, \ldots, |X_{n-1}|$. Thus, with probability one,

$$0 < V_1 < V_2 < \cdots < V_{n-1} = M.$$

Assume $\sum_{t=1}^{n-1} V_t > 4M$. This reasonable assumption,[1] along with the fact that $1 < x^2 < 4$, gives us that

$$\sum_{t=1}^{n-1} \left(\frac{x^2 V_t}{4M} \right)^2 < x^2 \sum_{t=1}^{n-1} \frac{V_t}{4M} < x^2 \left(\sum_{t=1}^{n-1} \frac{V_t}{4M} \right)^2,$$

which implies that

$$A = \left(\sum_{t=1}^{n-1} V_t \right)^2 - x^2 \sum_{t=1}^{n-1} V_t^2 > 0,$$

and if $A > 0$, then $a > 0$. Also, because $b^2 - 4ac$ is equal to

$$\frac{x^2}{(\sum_{t=2}^{n} |X_{t-1}|)^2} \left[\sum_{t=2}^{n} (X_t - \widetilde{\phi}_0 X_{t-1})^2 - x^2 r \sum_{t=2}^{n} (X_t - \widetilde{\phi}_1 X_{t-1})^2 \right]$$

and since, by definition,

$$\sum_{t=2}^{n} (X_t - \widetilde{\phi}_1 X_{t-1})^2 \leqslant \sum_{t=2}^{n} (X_t - \widetilde{\phi}_0 X_{t-1})^2,$$

then if $x^2 r < 1$ (i.e., $a > 0$), we have $b^2 - 4ac > 0$. Thus, once $\sum_{t=1}^{n-1} V_t > 4M$, our parabola opens upward and has real intercepts.

References

Ahtola, J., Tiao, G.C., 1984. Parameter inference for a nearly nonstationary first-order autoregressive model. Biometrika 71, 263–272.
Ali, M.M., 2002. Distribution of the least squares estimator in a first-order autoregressive model. Econometric Rev. 21, 89–119.
Andrews, D.W.K., 1993. Exactly median-unbiased estimation of first order autoregressive/unit root models. Econometrica 61, 139–165.
Basawa, I.V., Mallik, A.K., McCormick, W.P., Reeves, J.H., Taylor, R.L., 1991. Bootstrapping unstable first-order autoregressive processes. Ann. Statist. 19, 1098–1101.
Bentkus, V., Götze, F., 1996. The Berry–Esseen bound for student's statistic. Ann. Probab. 24, 491–503.
Carpenter, J., 1999. Test inversion bootstrap confidence intervals. J. R. Statist. Soc. B 61, 159–172.
Durrett, R., 1996. Probability: Theory and Examples. second ed. Duxbury, New York.

[1] For 10,000 simulations with $\{\varepsilon_t\} \sim t_3$, this assumption always held for ϕ between -1.5 and 1.5 once $n \geqslant 12$.

Efron, B., 1969. Student's t-test under symmetry conditions. J. Amer. Statist. Assoc. 64, 1278–1302.
Elliot, G., Stock, J.H., 2001. Confidence intervals for autoregressive coefficients near one. J. Econometrics 103, 155–181.
Fuller, W.A., 1996. Introduction to Statistical Time Series. second ed. Wiley, New York.
Hansen, B.E., 1999. The grid bootstrap and the autoregressive model. Rev. Econometrics Statist. 81, 594–607.
Lieberman, O., 1994. Saddlepoint approximation for the least squares estimator in first-order autoregression. Biometrika 81 (4), 807–811.
Mann, H.B., Wald, A., 1943. On the statistical treatment of linear stochastic difference equations. Econometrica 11, 173–220.
Marriott, F.H.C., Pope, J.A., 1954. Bias in the estimation of autocorrelations. Biometrika 41, 390–402.
Newbold, P., Agiakloglou, C., 1993. Bias in the sample autocorrelations of fractional noise. Biometrika 80, 698–702.
Perron, P.A., 1991. Continuous time approximation to the unstable first-order autoregressive process: the case without an intercept. Econometrica 59 (1), 211–236.
Phillips, P.C.B., 1978. Edgeworth and saddlepoint approximations in the first-order noncircular autoregression. Biometrika 65, 91–98.
Phillips, P.C.B., Park, J.Y., Chang, Y., 2004. Nonlinear instrumental variable estimation of an autoregression. J. Econometrics 118, 219–246.
Rao, M.M., 1978. Asymptotic distribution of an estimator of the boundary parameter of an unstable process. Ann. Statist. 6, 185–190.
Shaman, P., Stine, R.A., 1988. The bias of autoregressive coefficient estimators. J. Amer. Statist. Assoc. 83, 842–848.
So, B.S., Shin, D.W., 1999. Cauchy estimators for autoregressive processes. Econometric Theory 15, 165–176.
Wang, S., 1992. Tail probability approximations in the first-order noncircular autoregression. Biometrika 79, 431–434.
White, J.S., 1958. The limiting distribution of the serial correlation coefficient in the explosive case. Ann. Math. Statist. 29, 1188–1197.