Journal of Statistical Planning and Inference 140 (2010) 297 – 309

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Corrected confidence intervals for parameters in adaptive linear models

Shen-Chien Chen[a], Ruby C. Weng[b,*], Tzeeming Huang[b]

[a] *Chungyu Institute of Technology, Keelung 20103, Taiwan*
[b] *National Chengchi University, Taipei 11623, Taiwan*

ARTICLE INFO

Article history:
Received 6 December 2008
Received in revised form
10 July 2009
Accepted 10 July 2009
Available online 18 July 2009

Keywords:
Confidence sets
Threshold autoregressive model
Very weak expansions

ABSTRACT

Consider an adaptive linear model $y_t = x_t'\theta + \sigma e_t$, where $x_t = (x_{t1}, \ldots, x_{tp})'$ may depend on previous responses. Woodroofe and Coad [1999. Corrected confidence sets for sequentially designed experiments: examples. In: Ghosh, S. (Ed.), Multivariate Analysis, Design of Experiments, and Survey Sampling. Marcel Dekker, Inc., New York, pp. 135–161] derived very weak asymptotic expansions for the distributions of an appropriate pivotal quantity and constructed corrected confidence sets for θ, where the correction terms involve the limit of $\sum_{t=1}^{n} x_t x_t'/n$ (as n approaches infinity) and its derivatives with respect to θ. However, the analytic form of this limit and its derivatives may not be tractable in some models. This paper proposes a numerical method to approximate the correction terms. For the resulting approximate pivot, we show that under mild conditions the error induced by numerical approximation is $o_p(1/n)$. Then, we assess the accuracy of the proposed method by an autoregressive model and a threshold autoregressive model.

1. Introduction

Consider an adaptive linear model of the form

$$y_t = x_t'\theta + \sigma e_t, \quad t = 1, 2, \ldots, \tag{1}$$

where $e_1, e_2, \ldots$ are i.i.d. standard normal random variables and $\theta = (\theta_1, \ldots, \theta_p)'$ and $\sigma > 0$ are unknown parameters. Here "adaptive" means that $x_t = (x_{t1}, \ldots, x_{tp})'$ may depend on previous responses; that is $x_t = x_t(y_1, \ldots, y_{t-1})$. The above model is quite general and includes time series and control problems as in Lai and Wei (1982), adaptive biased coin designs as in Eisele (1994), among others.

An adaptive normal linear model with known variance was studied by Woodroofe (1989) and Woodroofe and Coad (1997). They used a Bayesian approach and Stein's (1981) identity to obtain asymptotic expansions for sampling distributions and construct corrected confidence sets for θ correct to order $o(1/n)$. The case of unknown σ was considered by Coad and Woodroofe (1998) and Woodroofe and Coad (1999), and approximations for confidence sets were evaluated for autoregressive (AR) processes, Ford–Silvey model, and certain clinical trial examples. It was shown that the maximum likelihood estimators may be severely biased in these models. This Bayesian approach starts with an approximate pivot, and employs Stein's identity to derive asymptotic expansions for the mean and variance corrections of the pivot. Then, it proceeds in the usual way to obtain the renormalized pivot, which is used to form corrected confidence sets. The correction terms have simple expressions, which involve the analytic

* Corresponding author.
E-mail address: chweng@nccu.edu.tw (R.C. Weng).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2009.07.014

forms of the limit of $\sum_{t=1}^{n} x_t x_t'/n$ (as n approaches infinity) and its derivatives with respect to the parameter θ. For example, in AR(2) models

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \sigma e_t, \quad t = 0, 1, \ldots, \tag{2}$$

the correction terms depend on the first two moments of the process and the derivatives of these moments, which are tractable. However, in some models the correction terms may not be available. Consider the following two-regime threshold autoregressive (TAR) model

$$y_t = \theta_1 y_{t-1}^+ + \theta_2 y_{t-1}^- + \sigma e_t, \quad t = 0, 1, \ldots, \tag{3}$$

where

$$y_{t-1}^+ = \begin{cases} y_{t-1} & \text{if } y_{t-1} > 0, \\ 0 & \text{if } y_{t-1} \le 0 \end{cases} \quad \text{and} \quad y_{t-1}^- = \begin{cases} 0 & \text{if } y_{t-1} > 0, \\ y_{t-1} & \text{if } y_{t-1} \le 0. \end{cases}$$

The TAR model was proposed by Tong (1983, 1990) to characterize certain nonlinear features of a process, and it became quite popular in the non-linear time series literature. In this model, we have

$$\frac{1}{n}\sum_{t=1}^{n} x_t x_t' = \begin{pmatrix} \frac{1}{n}\sum_{t=1}^{n}(y_{t-1}^+)^2 & 0 \\ 0 & \frac{1}{n}\sum_{t=1}^{n}(y_{t-1}^-)^2 \end{pmatrix} \to \begin{pmatrix} E_\theta(S^+)^2 & 0 \\ 0 & E_\theta(S^-)^2 \end{pmatrix}, \tag{4}$$

where S follows the stationary distribution of the process. However, it is known that the explicit analytic forms for the stationary distribution and moments of a simple TAR model are difficult to derive and are known only in certain cases where the autoregression function has special structures and the error term follows some specific distributions. For example, Anděl et al. (1984) studied model (3) with $\theta_1 = -\theta_2$, $\theta_1 \in (0, 1)$, and e_t follows N(0, 1), and Anděl and Bartoň (1986) and Loges (2004) considered the same model with Cauchy and Laplace distributions respectively. Consequently, for TAR models the corrections suggested by very weak type approximations cannot be obtained in the usual way. To address this problem, we propose to approximate the limit of $\sum_{t=1}^{n} x_t x_t'/n$ and its derivatives by combining the difference quotient method and Monte Carlo simulations. Then, we show that under mild conditions the obtained confidence intervals are accurate to order $o_p(1/n)$, where o_p is in the sense of (31).

We organize the remainder of this paper as follows. The next section gives brief review of very weak approximation for adaptive linear models. In Section 3 we describe our method and conduct error analysis. In Section 4 we assess the accuracy of the proposed method by simulation studies. Section 5 concludes the paper.

2. Review

It is known that the likelihood function is not affected by the adaptive nature of the design, so the maximum likelihood estimator of θ has the form

$$\hat{\theta}_n = \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} \left(\sum_{t=1}^{n} x_t y_t\right),$$

provided that $\sum_{t=1}^{n} x_t x_t'$ is positive definite. The usual estimator of σ^2 is

$$\hat{\sigma}_n^2 = \frac{\sum_{t=1}^{n}(y_t - x_t'\hat{\theta}_n)^2}{n-p}.$$

Let B_n be a $p \times p$ matrix for which

$$B_n B_n' = \sum_{t=1}^{n} x_t x_t', \tag{5}$$

and define

$$Z_n = \frac{1}{\sigma} B_n'(\theta - \hat{\theta}_n), \quad T_n = \frac{1}{\hat{\sigma}_n} B_n'(\theta - \hat{\theta}_n). \tag{6}$$

These are served as the first approximate pivots for known and unknown σ respectively. The variables Z_n and T_n have exactly p-variate standard normal and t distributions respectively, in the absence of an adaptive design. For the case of an adaptive design, the bias correction is needed. To describe the correction term, we first introduce some notations. Let $P_{\theta,\sigma}$ denote the probability distribution of the model, and $E_{\theta,\sigma}$ the expectation with respect to $P_{\theta,\sigma}$. Since $\hat{\theta}_n$ is invariant and B_n is equivalent, the distributions of Z_n and T_n do not depend on σ. So, without loss of generality we take $\sigma = 1$ for discussion. Hereafter σ may be suppressed; for example, we abbreviate $P_{\theta,1}$ and $E_{\theta,1}$ as P_θ and E_θ, respectively. Suppose that

$$Q_n \equiv \sqrt{n}B_n^{-1} \to Q_\theta \quad \text{in } P_\theta\text{-probability,} \tag{7}$$

where $Q_\theta = [q_{ij}(\theta): i,j = 1,\dots,p]$. Suppose that q_{ij} are differentiable with respect to θ, and let $Q_\theta^\# = [q_{ij}^\#(\theta) : i,j = 1,\dots,p]$ and $M_\theta = [m_{ij}(\theta) : i,j = 1,\dots,p]$ where

$$q_{ij}^\#(\theta) = \frac{\partial q_{ij}(\theta)}{\partial \theta_j} \tag{8}$$

and

$$m_{ij}(\theta) = \sum_{k=1}^{p}\sum_{l=1}^{p} \frac{\partial^2}{\partial\theta_k \partial\theta_l}[q_{ik}(\theta) q_{jl}(\theta)]. \tag{9}$$

Let Φ^p denote the standard p-variate normal distribution. For $h : \Re^p \to \Re$, write

$$\Phi^p h = \int h(w)\Phi^p(dw), \quad \Phi_1^p h = \int w h(w)\Phi^p(dw) \quad (p \times 1), \quad \Phi_2^p h = \frac{1}{2}\int (ww' - I_p) h(w)\Phi^p(dw) \quad (p \times p),$$

whenever the integrals exist. To find the mean and covariance of Z_n, note that if h is a function of quadratic growth, then

$$E_\theta[h(Z_n)] \approx \Phi^p h - \frac{1}{\sqrt{n}}(\Phi_1^p h)' Q_\theta^\# \mathbf{1} + \frac{1}{n}\mathrm{tr}[(\Phi_2^p h) M_\theta], \tag{10}$$

where $\mathbf{1} = (1,\dots,1)'$. Specializing (10) to $h(z) = z_i$ and $h(z) = z_i z_j$ gives approximate mean and covariance of Z_n:

$$E_\theta(Z_n) \approx -\frac{1}{\sqrt{n}} Q_\theta^\# \mathbf{1} = \mu_n(\theta), \quad \text{say} \tag{11}$$

and

$$E_\theta(Z_n Z_n') \approx I_p + \frac{1}{n} M_\theta. \tag{12}$$

Let $\Delta_\theta = [\delta_{ij}(\theta) : i,j = 1,\dots,p]$ with

$$\delta_{ij}(\theta) = \sum_{k=1}^{p}\sum_{l=1}^{p} \left(\frac{\partial q_{ik}}{\partial \theta_l}\right)\left(\frac{\partial q_{jl}}{\partial \theta_k}\right). \tag{13}$$

Let $\hat{\mu}_n = \mu_n(\hat{\theta}_n)$. For the case of known σ, it can be shown that

$$E_\theta[(Z_n - \hat{\mu}_n)(Z_n - \hat{\mu}_n)'] \approx I_p + \frac{\Delta_\theta}{n}.$$

Let $\Gamma_n = \Gamma_n(\theta)$ be a $p \times p$ matrix such that

$$\Gamma_n \Gamma_n' = I_p + \frac{\Delta_\theta}{n}. \tag{14}$$

Let $\hat{\Gamma}_n = \Gamma_n(\hat{\theta}_n)$ and define

$$Z_n^* = \hat{\Gamma}_n^{-1}(Z_n - \hat{\mu}_n).$$

Then, it can be shown that

$$E_\theta[h(Z_n^*)] \approx \Phi^p h. \tag{15}$$

For the case of unknown σ, we start with T_n and define the renormalized pivot as

$$T_n^* = \hat{\Gamma}_n^{-1}(T_n - \hat{\mu}_n). \tag{16}$$

If h is a function of quadratic growth, then

$$E_\theta[h(T_n^*)] \approx \Phi^p h + \frac{\Phi_4^p h}{n}, \tag{17}$$

where

$$\Phi_4^p h = \int_{\Re^p} \left\{ \frac{1}{4}[\|z\|^2 - p]^2 - \frac{1}{2}p \right\} h(z)\Phi^p\{dz\}.$$

The convergence of (15) and (17) are in the very weak sense of Woodroofe (1986, 1989); that is,

$$\int_{\Omega} [P_\theta\{Z_n^* \in B\} - \Phi^p(B)]\xi(\theta)\,d\theta = \mathrm{o}\left(\frac{1}{n}\right) \quad \text{and} \quad \int_{\Omega} [P_\theta\{T_n^* \in B\} - G_n^p(B)]\xi(\theta)\,d\theta = \mathrm{o}\left(\frac{1}{n}\right) \tag{18}$$

uniformly with respect to Borel sets $B \subseteq \Re^p$ for all twice continuously differentiable densities ξ with compact convex support, where G_n^p is the spherically symmetric p-variate t distribution with n degrees of freedom. Woodroofe (1989) writes relation (18) as

$$P_\theta\{T_n^* \in B\} = G_n^p(B) + \mathrm{o}\left(\frac{1}{n}\right) \tag{19}$$

very weakly, and argues that (19) is strong enough to support a frequentist interpretation for confidence intervals.

It is easy to use (19) to form corrected confidence sets. By (5) we can take B_n as a lower triangular matrix, and for convenience we take $\hat{\Gamma}_n$ as an upper triangular matrix satisfying $\hat{\Gamma}_n\hat{\Gamma}_n' = I_p + \Delta_{\hat{\theta}_n}/n$. Since Δ_θ is symmetric, $\hat{\Gamma}_n \approx I_p + \Delta_{\hat{\theta}_n}/(2n)$. Let $\hat{\mu}_{ni}$ and T_{ni}^* be the i-th components of $\hat{\mu}_n$ and T_n^* respectively, $\hat{\delta}_{n,ij}$ be the (i,j)-th entry of $\Delta_{\hat{\theta}_n}$, and b_n be the lower right-hand entry in B_n. Then, an asymptotic level γ confidence interval for θ_p is $\{|T_{np}^*| \le c_n\}$, where c_n is the $100(1+\gamma)/2$ quantile of the standard univariate t-distribution with n degrees of freedom; that is,

$$\hat{\theta}_{np} + \frac{\hat{\sigma}_n}{b_n}\hat{\mu}_{np} \pm \frac{\hat{\sigma}_n}{b_n}\left(1 + \frac{\hat{\delta}_{n,pp}}{2n}\right) \times c_n. \tag{20}$$

The above results can be found in Sections 2–4 of Woodroofe and Coad (1999).

3. Proposed method

Recall from (8), (14), and (16) that the approximate mean and variance of T_n involve $q_{ij}^{\#}(\hat{\theta}_n)$ and $\delta_{ij}(\hat{\theta}_n)$ defined in (8) and (13), which are the partial derivatives of q_{ij} with respect to θ evaluated at $\hat{\theta}_n$. Throughout this section we denote $\partial q_{ij}(\theta)/\partial\theta_k$ as $q_{ij,k}(\theta)$. To tackle the problem that the analytic forms of $q_{ij,k}(\theta)$ are not tractable, we propose to approximate $q_{ij,k}(\hat{\theta}_n)$ by numerical methods. In Section 3.1 we describe the approximation procedure and in Section 3.2 we show that the proposed pivot $T_n^\dagger$ in (27) differs from T_n^* by $\mathrm{o_p}(1/n)$.

3.1. Approximation procedure

Our approximation consists of two parts: difference quotient and Monte Carlo simulations. First, the difference quotient method suggests

$$q_{ij,k}(\hat{\theta}_n) \equiv \frac{\partial q_{ij}}{\partial\theta_k}(\hat{\theta}_n) \approx \frac{q_{ij}(\hat{\theta}_n + \eta e_k) - q_{ij}(\hat{\theta}_n)}{\eta}, \tag{21}$$

where η is a small positive number and the set $\{e_j : j = 1, \ldots, p\}$ forms an orthonormal basis of $\Re^p$. Note that by the Mean Value Theorem the right side of (21) can be expressed as

$$\frac{q_{ij}(\hat{\theta}_n + \eta e_k) - q_{ij}(\hat{\theta}_n)}{\eta} = \frac{\partial q_{ij}}{\partial\theta_k}(\theta_n^*) \equiv q_{ij,k}(\theta_n^*), \tag{22}$$

where θ_n^* lies between $\hat{\theta}_n$ and $\hat{\theta}_n + \eta e_k$. Next, consider $q_{ij}(\hat{\theta}_n + \eta e_k)$ and $q_{ij}(\hat{\theta}_n)$ in the numerator of (21). Suppose that

$$\frac{1}{n}\sum_{t=1}^{n} x_t x_t' \to A(\theta) = [a_{ij}(\theta) : i,j = 1, \ldots, p] \tag{23}$$

in P_θ-probability. Define the $p(p+1)/2$-dimensional vectors $a(\theta) = \{a_{lr}(\theta) : 1 \le l \le r \le p\}$ and $a^{(m)}(\theta) = \{a_{lr}^{(m)}(\theta) : 1 \le l \le r \le p\}$, where $\theta \in \Omega$ and $a_{lr}^{(m)}$ are empirical approximations of a_{lr} based on a Monte Carlo simulated sample of size m. By the relation of B_n and Q_θ in (7) and the definition of B_n in (5), it is not difficult to see that each entry of B_n is a function of $\{\sum_{k=1}^{n} x_{ki}x_{kj} : 1 \le i \le j \le p\}$ and $q_{ij}(\theta)$'s are functions of $a(\theta)$; that is,

$$q_{ij}(\theta) = h_{ij}(a(\theta)) \tag{24}$$

for some known functions $h_{ij}:\Re^{p(p+1)/2}\to\Re$. We have noted that the explicit analytic form for the $a(\theta)$ may not be available. Therefore, we propose to approximate $a(\theta)$ by a Monte Carlo simulation $a^{(m)}(\theta)$, and approximate $q_{ij}(\theta)=h_{ij}(a(\theta))$ by $h_{ij}(a^{(m)}(\theta))$. Then, by (21) and (22), the resulting estimate of $q_{ij,k}(\hat\theta_n)$ is

$$\tilde q_{ij,k}=\frac{h_{ij}(a^{(m)}(\hat\theta_n+\eta e_k))-h_{ij}(a^{(m)}(\hat\theta_n))}{\eta}. \tag{25}$$

Note that $\tilde q_{ij,k}$ depends on the choice of η and m, but the dependence is suppressed in the notation for convenience. Denote the corresponding approximations to $Q_\theta^\#$, Δ_θ, μ_n and Γ_n as $\tilde Q_\theta^\#$, $\tilde\Delta_\theta$, $\tilde\mu_n$ and $\tilde\Gamma_n$. So, we have

$$\tilde\mu_n=-\frac{1}{\sqrt n}\tilde Q_\theta^\#\mathbf{1}\quad\text{and}\quad\tilde\Gamma_n\tilde\Gamma_n'=I_p+\frac{\tilde\Delta_\theta}{n}. \tag{26}$$

From these we proceed to obtain the renormalized pivot as

$$T_n^\dagger=\tilde\Gamma_n^{-1}(T_n-\tilde\mu_n). \tag{27}$$

As an illustration of the above procedure, we consider the approximation of $q_{22,1}(\hat\theta_n)$ in the TAR model (3). Let S denote a random variable that follows the stationary distribution of the model. Then, by (4) and (23) we have

$$\frac1n\sum_{t=1}^n x_tx_t'\to\begin{pmatrix}E_\theta(S^+)^2 & 0\\ 0 & E_\theta(S^-)^2\end{pmatrix}=\begin{pmatrix}a_{11}(\theta) & 0\\ 0 & a_{22}(\theta)\end{pmatrix}=A(\theta),$$

and by (5) and (7) the inverse of $A(\theta)$ is exactly $Q_\theta'Q_\theta$. So, Q_θ can be expressed as

$$Q_\theta=\begin{pmatrix}(E_\theta(S^+)^2)^{-1/2} & 0\\ 0 & (E_\theta(S^-)^2)^{-1/2}\end{pmatrix}, \tag{28}$$

and we can write $q_{22}(\theta)=[a_{22}(\theta)]^{-1/2}$, where $a_{22}(\theta)=E_\theta(S^-)^2$. To approximate $q_{22,1}(\hat\theta_n)$, first by the difference quotient in (21) we have

$$q_{22,1}(\hat\theta_n)\equiv\frac{\partial q_{22}(\hat\theta_n)}{\partial\theta_1}\approx\frac{q_{22}(\hat\theta_n+(\eta,0)')-q_{22}(\hat\theta_n)}{\eta}=\frac{[a_{22}(\hat\theta_n+(\eta,0)')]^{-1/2}-[a_{22}(\hat\theta_n)]^{-1/2}}{\eta},$$

where

$$a_{22}(\hat\theta_n+(\eta,0)')=E_{\hat\theta_n+(\eta,0)'}(S^-)^2\text{ and }a_{22}(\hat\theta_n)=E_{\hat\theta_n}(S^-)^2.$$

Next, we simulate two stationary series W_{1i} and W_{2i} of length m with underlying parameters $\hat\theta_n$ and $\hat\theta_n+(\eta,0)'$, respectively, and approximate $a_{22}(\hat\theta_n+(\eta,0)')$ and $a_{22}(\hat\theta_n)$ by the empirical moments $a_{22}^{(m)}(\hat\theta_n+(\eta,0)')\equiv\sum_{i=1}^m(W_{2i}^-)^2/m$ and $a_{22}^{(m)}(\hat\theta_n)\equiv\sum_{i=1}^m(W_{1i}^-)^2/m$, respectively. Then, the resulting approximate to $q_{22,1}(\hat\theta_n)$ is

$$\tilde q_{22,1}=\frac{\left(\frac1m\sum_{i=1}^m(W_{2i}^-)^2\right)^{-1/2}-\left(\frac1m\sum_{i=1}^m(W_{1i}^-)^2\right)^{-1/2}}{\eta}. \tag{29}$$

In practice, the $\tilde q_{ij,k}$ in (25) can be obtained without knowing the function form of h_{ij}. To see how, first observe that from (5), (7), and (23) we have

$$A^{-1}(\theta)=Q_\theta'Q_\theta. \tag{30}$$

Next, define the $p\times p$ matrix $A^{(m)}(\hat\varphi_n)=[a_{ij}^{(m)}(\hat\varphi_n):\ i,j=1,\dots,p]$, where $\hat\varphi_n=\hat\theta_n$ or $\hat\theta_n+\eta e_k$, $k=1,\dots,p$. So, $Q_{\hat\varphi_n}$ satisfies $A^{-1}(\hat\varphi_n)=Q_{\hat\varphi_n}'Q_{\hat\varphi_n}$ and it can be estimated by the Cholesky decomposition of $[A^{(m)}(\hat\varphi_n)]^{-1}$. Denote the obtained estimates as $Q_{\hat\varphi_n}^{(m)}$. Then, from these $Q_{\hat\varphi_n}^{(m)}$ matrices we obtain $q_{ij,k}(\hat\theta_n)$ by difference quotient method.

3.2. Error analysis

Throughout this section, we assume that the true parameter θ_0 is an interior point of the parameter space Ω. The $o_p(1)$, $O_p(1)$ and $\xrightarrow{p}$ (convergence in p-probability) hereafter are with respect to some probability measure p, where p can be P_{θ_0}, $\bar P$ or $\bar P_\theta$ (defined before Lemma 3.2). Next, let $B_r(w;d)$ denote the r-dimensional open ball with radius d centered at w, and let $\bar B_r(w;d)$

denote the closure of it. Now, we will analyze the errors induced from approximating $q_{ij,k}(\hat{\theta}_n)$ by $\tilde{q}_{ij,k}$, and we will show that $T_n^\dagger$ defined in (27) differs from T_n^* in (16) by $\mathrm{o_p}(1/n)$; that is,

$$T_n^\dagger = T_n^* + \mathrm{o_p}(n^{-1}). \tag{31}$$

To begin, note that by adding and subtracting $q_{ij,k}(\theta_n^*)$ defined in (22) to the difference $q_{ij,k}(\hat{\theta}_n) - \tilde{q}_{ij,k}$ yields

$$q_{ij,k}(\hat{\theta}_n) - \tilde{q}_{ij,k} = [q_{ij,k}(\hat{\theta}_n) - q_{ij,k}(\theta_n^*)] + [q_{ij,k}(\theta_n^*) - \tilde{q}_{ij,k}] = (I) + (II), \quad \text{say}, \tag{32}$$

where the error (I) is from approximating the derivative by difference quotient, while (II) is from approximating the moments $\{a_{lr}\}$ by Monte Carlo simulations $\{a_{lr}^{(m)}\}$.

The following lemma deals with the order of (I).

Lemma 3.1. *Suppose that $\hat{\theta}_n \xrightarrow{p} \theta_0$ and that $\partial^2 q_{ij}/\partial\theta_k^2$ is continuous in Ω. Then, $(I) = \mathrm{O_p}(\eta)$.*

Proof. To start, note that by (22) $\hat{\theta}_n$ and θ_n^* differ only in the k th component, and that $\|\hat{\theta}_n - \theta_n^*\| = |\hat{\theta}_{nk} - \theta_{nk}^*| \le \eta$. Next, write

$$(I) = \frac{\partial q_{ij}(\hat{\theta}_n)}{\partial\theta_k} - \frac{\partial q_{ij}(\theta_n^*)}{\partial\theta_k} = \frac{\partial^2 q_{ij}(\zeta)}{\partial\theta_k^2}(\hat{\theta}_{nk} - \theta_{nk}^*), \tag{33}$$

where (33) follows from the Mean Value Theorem, and ζ lies between $\hat{\theta}_n$ and θ_n^*. Since θ_0 is an interior point of Ω, it implies that there exist $\varepsilon > \eta > 0$ such that the ball $\bar{B}_p(\theta_0;\varepsilon) \equiv \{w : \|w - \theta_0\| \le \varepsilon\}$ lies in the parameter space Ω. So,

$$\|\zeta - \theta_0\| \le \|\zeta - \hat{\theta}_n\| + \|\hat{\theta}_n - \theta_0\| \le \|\hat{\theta}_n - \theta_n^*\| + \|\hat{\theta}_n - \theta_0\| \le \eta + \|\hat{\theta}_n - \theta_0\| < \varepsilon,$$

provided that n is large enough. Then,

$$(33) \le \sup_{w:\|w-\theta_0\|\le\varepsilon} \left[\frac{\partial^2 q_{ij}(w)}{\partial\theta_k^2}\right]\eta = \mathrm{O_p}(\eta), \tag{34}$$

by continuity of $\partial^2 q_{ij}/\partial\theta_k^2$ and compactness of $\bar{B}_p(\theta_0;\varepsilon)$. $\square$

For (II), by (22), (24), and (25), we have

$$(II) = \frac{\partial q_{ij}}{\partial\theta_k}(\theta_n^*) - \tilde{q}_{ij,k} = \frac{1}{\eta}\{[h_{ij}(a(\hat{\theta}_n + \eta e_k)) - h_{ij}(a^{(m)}(\hat{\theta}_n + \eta e_k))] - [h_{ij}(a(\hat{\theta}_n)) - h_{ij}(a^{(m)}(\hat{\theta}_n))]\}. \tag{35}$$

In the next lemma we consider the order of $a_{lr}(\hat{\varphi}_n) - a_{lr}^{(m)}(\hat{\varphi}_n)$, where $\hat{\varphi}_n = \hat{\theta}_n$ or $\hat{\theta}_n + \eta e_k$ for $k = 1, \ldots, p$. Since $a_{lr}(\varphi)$ is some expectation and $a_{lr}^{(m)}(\varphi)$ is an empirical approximation, we can write

$$a_{lr}^{(m)}(\varphi) - a_{lr}(\varphi) = \frac{1}{m}\sum_{i=1}^m W_{i,\varphi} - EW_\varphi,$$

where $EW_\varphi = a_{lr}(\varphi)$ and $\{W_{i,\varphi}\}_{i=1}^m$ is a simulated sample of W_φ. Recall in Section 2 that P_θ denotes the probability distribution of data $\{(x_t, y_t)\}_{t=1}^n$ from the model (1). In the lemma below we let $\bar{P}$ denote the probability distribution of $\{W_{i,\varphi}\}_{i=1}^m$ and $\bar{P}_\theta$ denote the joint probability distribution of $\{(x_t, y_t)\}_{t=1}^n$ and $\{W_{i,\hat{\varphi}_n}\}_{i=1}^m$.

Lemma 3.2. *Suppose that $\eta = \eta_n \to 0$, $\hat{\varphi}_n \xrightarrow{p} \theta_0$, and there exist $\delta > 0$, $\alpha > 0$, and $m_0 > 0$ such that*

$$\sup_{\varphi\in\bar{B}_p(\theta_0;\delta)} \bar{E}\left(\frac{1}{m}\sum_{i=1}^m W_{i,\varphi} - EW_\varphi\right)^2 \le \frac{\alpha_0}{m} \tag{36}$$

for all $m > m_0$. Then,

$$\frac{1}{m}\sum_{i=1}^m W_{i,\hat{\varphi}_n} - EW_{\hat{\varphi}_n} = \mathrm{O_p}(1/\sqrt{m}).$$

Proof. Given $\varepsilon > 0$, it suffices to show that there exist some $c_1 > 0$ and $n_0 > 0$ such that

$$\bar{P}_{\theta_0}\left(\left|\frac{1}{m}\sum_{i=1}^{m} W_{i,\hat{\varphi}_n} - EW_{\hat{\varphi}_n}\right| > \frac{c_1}{\sqrt{m}}\right) \le \varepsilon$$

for all $n > n_0$. To begin, choose n_0 such that $P_{\theta_0}(\|\hat{\varphi}_n - \theta_0\| > \delta) \le \varepsilon/2\ \forall n > n_0$, and let $c_1 \ge \sqrt{2\alpha_0/\varepsilon}$. Then, for $n > n_0$

$$\begin{aligned}\bar{P}_{\theta_0}\left(\left|\frac{1}{m}\sum_{i=1}^{m} W_{i,\hat{\varphi}_n} - EW_{\hat{\varphi}_n}\right| > \frac{c_1}{\sqrt{m}}\right) &= E_{\theta_0}\left[E\left(1_{\left\{\left|\frac{1}{m}\sum_{i=1}^{m} W_{i,\hat{\varphi}_n} - EW_{\hat{\varphi}_n}\right| > c_1/\sqrt{m}\right\}} | \hat{\varphi}_n\right)\right] \\ &\le P_{\theta_0}(\|\hat{\varphi}_n - \theta_0\| > \delta) + \int_{\bar{B}_p(\theta_0;\delta)} \bar{P}\left(\left|\frac{1}{m}\sum_{i=1}^{m} W_{i,\varphi} - EW_{\varphi}\right| > \frac{c_1}{\sqrt{m}}\right) dF_{\hat{\varphi}_n}(\varphi), \end{aligned} \tag{37}$$

where the first expectation in (37) is with respect to the distribution of $\hat{\theta}_n$ when the underlying true parameter is θ_0, the second expectation in (37) is with respect to the conditional distribution of $\{W_{i,\hat{\varphi}_n}\}_{i=1}^m$ given the value of $\hat{\varphi}_n$, and $F_{\hat{\varphi}_n}$ denotes the probability distribution of $\hat{\varphi}_n$. By (36) and Markov inequality we have

$$\sup_{\varphi \in \bar{B}_p(\theta_0;\delta)} \bar{P}\left(\left|\frac{1}{m}\sum_{i=1}^{m} W_{i,\varphi} - EW_{\varphi}\right| > \frac{c_1}{\sqrt{m}}\right) \le \frac{m}{c_1^2}\cdot\frac{\alpha_0}{m} \le \frac{\varepsilon}{2}.$$

So, (37) $\le \varepsilon/2 + \varepsilon/2 = \varepsilon$. This completes the proof. □

Lemma 3.3. *Suppose that the conditions in Lemma 3.2 hold, and that h_{ij} is continuously differentiable with respect to a_{lr} over the compact ball $\bar{B}_{p(p+1)/2}(a(\theta_0);\varepsilon)$ for some $\varepsilon > 0$. Then,* (II) $= O_p(1/(\eta\sqrt{m}))$.

The proof follows easily by first writing

$$h_{ij}(a(\hat{\varphi}_n)) - h_{ij}(a^{(m)}(\hat{\varphi}_n)) = \sum_{l \le r} \frac{\partial h_{ij}(\xi)}{\partial a_{lr}}[a_{lr}(\hat{\varphi}_n) - a_{lr}^{(m)}(\hat{\varphi}_n)],$$

where $\xi = \{\xi_{lr}, 1 \le l \le r \le p\}$ lies between the line segment joining $a(\hat{\varphi}_n)$ and $a^{(m)}(\hat{\varphi}_n)$. Then, the result follows by Lemma 3.2 and the assumption that h_{ij} is continuously differentiable.

To illustrate the verification of these lemmas, we consider approximating $q_{22,1}(\hat{\theta}_n)$ by $\tilde{q}_{22,1}$ in the TAR model (3). It is not difficult to see that q_{22} is twice continuously differentiable. Therefore, Lemma 3.1 is satisfied. Next, consider $q_{22,1}(\theta_n^*) - \tilde{q}_{22,1}$. From (28) and (35),

$$\begin{aligned} q_{22,1}(\theta_n^*) - \tilde{q}_{22,1} &= \frac{1}{\eta}\{[h_{22}(a(\hat{\theta}_n + \eta e_1)) - h_{22}(a^{(m)}(\hat{\theta}_n + \eta e_1))] - [h_{22}(a(\hat{\theta}_n)) - h_{22}(a^{(m)}(\hat{\theta}_n))]\} \\ &= \frac{1}{\eta}\left\{\left\{[a_{22}(\hat{\theta}_n + \eta e_1)]^{-1/2} - [a_{22}^{(m)}(\hat{\theta}_n + \eta e_1)]^{-1/2}\right\} - \left\{[a_{22}(\hat{\theta}_n)]^{-1/2} - [a_{22}^{(m)}(\hat{\theta}_n)]^{-1/2}\right\}\right\}, \end{aligned} \tag{38}$$

where the last line follows because

$$h_{22}(a) = h_{22}(a_{11}, a_{12}, a_{22}) = a_{22}^{-1/2}(\theta).$$

By (3) and (28), it is easy to see that there exists $\nu > 0$ such that $a_{22}(\theta_0) = E_{\theta_0}(S^-)^2 > \nu$, and that $h_{22}(a)$ is continuously differentiable on $\bar{B}_3(a(\theta_0);\varepsilon)$ for $0 < \varepsilon < \nu$. Finally, if the expectation $\bar{E}(\sum_{i=1}^m W_{i,\varphi}/m - EW_\varphi)^2$ in (36) of Lemma 3.2 is bounded by $C(\varphi)/m$, where $C(\varphi)$ is some continuous function of φ, then the supremum of $C(\varphi)$ over a compact region $\bar{B}_p(\theta_0;\delta)$ is bounded by some $\alpha_0 > 0$ and we have (36).

Now, from (27) we can write

$$T_n^\dagger = \tilde{\Gamma}_n^{-1}(T_n - \tilde{\mu}_n) = \tilde{\Gamma}_n^{-1}\hat{\Gamma}_n T_n^* + \tilde{\Gamma}_n^{-1}(\hat{\mu}_n - \tilde{\mu}_n).$$

So,

$$T_n^\dagger - T_n^* = (\tilde{\Gamma}_n^{-1}\hat{\Gamma}_n - I_p)T_n^* + \tilde{\Gamma}_n^{-1}(\hat{\mu}_n - \tilde{\mu}_n).$$

By (8), (11), and (26),

$$|\hat{\mu}_{n,i} - \tilde{\mu}_{n,i}| = \frac{1}{\sqrt{n}}\left|\sum_{j=1}^{p}(q_{ij}^{\#}(\hat{\theta}_n) - \tilde{q}_{ij,j})\right|,$$

and by (32) and Lemmas 3.1–3.3, we have

$$|q^{\#}_{ij}(\hat{\theta}_n) - \tilde{q}_{ij,j})| = \mathrm{O_p}(\eta) + \mathrm{O_p}\left(\frac{1}{\eta\sqrt{m}}\right).$$

Similarly, we can show that

$$\|\hat{\Gamma}_n - \tilde{\Gamma}_n\| = \frac{1}{2n}\|\Delta_{\hat{\theta}_n} - \tilde{\Delta}_\theta\| = \frac{1}{n}\left[\mathrm{O_p}(\eta) + \mathrm{O_p}\left(\frac{1}{\eta\sqrt{m}}\right)\right].$$

Then, we have the following result.

Proposition 3.4. *Suppose that the conditions in Lemmas* 3.1–3.3 *hold and that* η *and* m *are chosen so that* $\eta = \mathrm{O}(1/n)$ *and* $m = \mathrm{O}(n^4)$. *Then*, (31) *holds*.

Proof. Since $\eta = \mathrm{O}(1/n)$ and $m = \mathrm{O}(n^4)$, we have

$$\|\hat{\mu}_n - \tilde{\mu}_n\| = \frac{1}{\sqrt{n}}[\mathrm{O_p}(\eta) + \mathrm{O_p}(1/(\eta\sqrt{m}))] = \mathrm{o_p}(1/n), \quad \|\hat{\Gamma}_n - \tilde{\Gamma}_n\| = \frac{1}{n}[\mathrm{O_p}(\eta) + \mathrm{O_p}(1/(\eta\sqrt{m}))] = \mathrm{o_p}(1/n).$$

Next (26) implies $\tilde{\Gamma}_n = \mathrm{O_p}(1)$. Hence $\tilde{\Gamma}_n^{-1}\hat{\Gamma}_n - I_p = \tilde{\Gamma}_n^{-1}(\hat{\Gamma}_n - \tilde{\Gamma}_n) = \mathrm{o_p}(1/n)$, and $\tilde{\Gamma}_n^{-1}(\hat{\mu}_n - \tilde{\mu}_n) = \mathrm{o_p}(1/n)$. So, $T^{\dagger}_n - T^*_n = (\tilde{\Gamma}_n^{-1}\hat{\Gamma}_n - I_p)T^*_n + \tilde{\Gamma}_n^{-1}(\hat{\mu}_n - \tilde{\mu}_n) = \mathrm{o_p}(1/n)$. □

Recall that T^*_n is asymptotically t_n to order $\mathrm{o}(1/n)$ in the sense of (18) and that by (31) $T^{\dagger}_n$ differs from T^*_n by order $\mathrm{o_p}(1/n)$. So, similar to (20), an asymptotic level γ confidence interval for θ_p is $\{|T^{\dagger}_{np}| \le c_n\}$, where c_n is the $100(1+\gamma)/2$ quantile of the standard univariate t-distribution with n degrees of freedom; that is,

$$\hat{\theta}_{np} + \frac{\hat{\sigma}_n}{b_n}\tilde{\mu}_{np} \pm \frac{\hat{\sigma}_n}{b_n}\left(1 + \frac{\tilde{\delta}_{n,pp}}{2n}\right) \times c_n. \tag{39}$$

Note that the above interval differs from (20) derived by T^*_n only in the estimates of μ_{np} and δ_{pp}.

4. Experiments

In this section we assess the accuracy of the proposed method for an AR(2) example and a TAR(1) model. The AR(2) model has been studied by Woodroofe and Coad (1999) and Weng and Woodroofe (2006). We include this example to see how close the proposed numerical approach is to the analytic method by comparing the coverage probabilities using the T^*_n in (16) and the $T^{\dagger}_n$ in (27). For the TAR(1) model, T^*_n cannot be used. So, we compare $T^{\dagger}_n$ with bootstrap methods.

4.1. An AR(2) example

Consider the AR(2) model in (2). The parameter space Ω is determined by the inequalities: $\theta_1 + \theta_2 < 1$, $\theta_1 - \theta_2 > -1$, and $\theta_2 > -1$; see, for example, Brockwell and Davis [3, Chapter 8]. For corrected confidence sets of parameters in AR(2) using very weak type approximation, Woodroofe and Coad (1999) studied the case with fixed initial values $y_0 = y_{-1} = 0$, while Weng and Woodroofe (2006) considered the stationary case, where y_0 and y_{-1} were assumed to follow the limiting distribution of the process. The former paper showed that for $n = 25$ and 50, the simulated and nominal values of the coverage probability agree well, except for θ_2 near the vertices of the triangle Ω; the latter showed that for $n = 10, 20, 50$, the simulated coverage probabilities and tail probabilities of renormalized approximate pivot are much closer to the nominal values than the uncorrected one. Here we assume $y_0 = y_{-1} = 0$. Clearly, this model is of the form of (1) with $x_t = (y_{t-1}, y_{t-2})'$, and the limit of $\sum_{t=1}^n x_t x_t'/n$ can be easily obtained. Below we discuss confidence intervals for θ_2 only. The treatment for θ_1 is similar and hence omitted.

By (23), the $A(\theta)$ matrix in this model is

$$\frac{\sum_{t=1}^n x_t x_t'}{n} \to A(\theta) = \begin{pmatrix} E_\theta W_1^2 & E_\theta W_1 W_2 \\ E_\theta W_1 W_2 & E_\theta W_2^2 \end{pmatrix}, \tag{40}$$

where $\{W_t\}_{t=-\infty}^{\infty}$ follows the stationary AR(2) process (2), and the analytic form of $A(\theta)$ is easy to derive. Let Q_θ be a lower triangular matrix satisfying $Q_\theta' Q_\theta = A^{-1}(\theta)$ as in (30), then

$$Q_\theta = \begin{pmatrix} \sqrt{1 - \theta_2^2 - \dfrac{\theta_1^2(1+\theta_2)^2}{1-\theta_2^2}} & 0 \\ -\dfrac{\theta_1(1+\theta_2)}{\sqrt{1-\theta_2^2}} & \sqrt{1-\theta_2^2} \end{pmatrix}.$$

Table 1

(a) AR(2) model ($n = 30$, replicates $= 10,000$) (nominal Coverage $= 95\%$; $c_n = 2.042$; $\pm$ is the range within 1.96 standard deviations); (b) AR(2) model ($n = 50$, replicates $= 10,000$) (nominal Coverage $= 95\%$; $c_n = 2.008$; $\pm$ is the range within 1.96 standard deviations).

$(\theta_1\ \theta_2)$	$E_\theta(T_{n2})$ $E_\theta(T^*_{n2})$ $\bar{E}_\theta(T^\dagger_{n2})$	$E_\theta(T^2_{n2})$ $E_\theta(T^{*2}_{n2})$ $\bar{E}_\theta(T^{\dagger 2}_{n2})$	$P_\theta(T_{n2} \geq c_n)$ $P_\theta(T^*_{n2} \geq c_n)$ $\bar{P}_\theta(T^\dagger_{n2} \geq c_n)$	$P_\theta(T_{n2} \leq -c_n)$ $P_\theta(T^*_{n2} \leq -c_n)$ $\bar{P}_\theta(T^\dagger_{n2} \leq -c_n)$	$P_\theta(\lvert T_{n2}\rvert \leq c_n)$ $P_\theta(\lvert T^*_{n2}\rvert \leq c_n)$ $\bar{P}_\theta(\lvert T^\dagger_{n2}\rvert \leq c_n)$
(a)					
(0.0 −0.5)	0.014	0.979	0.021	0.021	0.958
	0.013	1.089	0.026	0.026	0.948
	0.013	1.088	0.026	0.026	0.948
(0.0 0.0)	0.180	1.018	0.033	0.011	0.956
	0.013	1.085	0.029	0.022	0.949
	0.013	1.085	0.028	0.021	0.950
(0.0 0.5)	0.375	1.106	0.048	0.007	0.945
	0.009	1.068	0.029	0.019	0.953
	0.010	1.067	0.029	0.018	0.953
(0.5 −0.5)	0.016	0.970	0.020	0.019	0.961
	0.014	1.078	0.024	0.024	0.952
	0.014	1.077	0.024	0.024	0.952
(0.5 −0.2)	0.114	0.972	0.027	0.014	0.959
	0.011	1.058	0.027	0.023	0.950
	0.011	1.057	0.027	0.023	0.950
(0.5 0.0)	0.177	0.983	0.031	0.011	0.958
	0.010	1.048	0.027	0.021	0.952
	0.010	1.047	0.027	0.021	0.952
$\pm$	± 0.020	1.07 ± 0.032	0.025 ± 0.003	0.025 ± 0.003	0.95 ± 0.004
(b)					
(0.0 −0.5)	0.017	0.986	0.023	0.022	0.955
	0.017	1.053	0.027	0.025	0.949
	0.016	1.052	0.027	0.025	0.949
(0.0 0.0)	0.141	1.007	0.030	0.016	0.954
	0.007	1.046	0.026	0.026	0.948
	0.007	1.046	0.026	0.025	0.949
(0.0 0.5)	0.294	1.062	0.042	0.012	0.947
	−0.006	1.037	0.026	0.024	0.951
	−0.004	1.037	0.026	0.024	0.951
(0.5 −0.5)	0.001	0.977	0.021	0.021	0.959
	−0.000	1.042	0.024	0.024	0.952
	−0.001	1.042	0.024	0.024	0.952
(0.5 −0.2)	0.076	0.990	0.025	0.020	0.955
	−0.007	1.045	0.024	0.028	0.948
	−0.007	1.045	0.024	0.028	0.948
(0.5 0.0)	0.127	0.999	0.027	0.018	0.955
	−0.008	1.042	0.024	0.027	0.949
	−0.008	1.042	0.024	0.027	0.949
$\pm$	± 0.020	1.04 ± 0.030	0.025 ± 0.003	0.025 ± 0.003	0.95 ± 0.004

From this, the earlier two papers derived $\partial q_{ij}/\partial \theta_k$ analytically, and from which the mean and variance corrections are readily available.

To implement the proposed numerical method in the present paper, for each of the parameter values in $\{\hat{\theta}_n, \hat{\theta}_n + \eta e_1, \hat{\theta}_n + \eta e_2\}$, we generate an AR(2) series of size m, and use simulated moments to approximate each entry in the matrices $A(\hat{\theta}_n)$ and $A(\hat{\theta}_n + \eta e_k)$, $k = 1, 2$. Then, the derivatives $q_{ij,k}(\hat{\theta}_n)$ are approximated as described following Eq. (30).

The selected slope parameters are the same as those in Weng and Woodroofe (2006). The values of η in (21) and m, the sample size of Monte Carlo simulation, should be determined. In the experiment, we chose $(\eta, m) = (0.001, 1000)$ and $(0.0005, 5000)$. As the results are close, we only report the case $(\eta, m) = (0.001, 1000)$. Tables 1a and b show the simulated values of $P_\theta(T^*_{n2} \geq c_n)$, $P_\theta(T^*_{n2} \leq -c_n)$ and $P_\theta(|T^*_{n2}| \leq c_n)$, for $\sigma = 1$ and $n = 30$ and 50; and similarly for $T^\dagger_{n2}$. Here c_n is the 2.5th percentile of the standard univariate t-distribution with n degrees of freedom. The notation $\pm$ in the last row indicates 1.96 standard deviations; for example, for $n = 50$, ± 0.020 is obtained by $E(t_{50}) \pm 1.96 \times [\mathrm{Var}(t_{50})/10,000]^{1/2}$, 1.25 ± 0.042 is by $E(t^2_{50}) \pm 1.96 \times [\mathrm{Var}(t^2_{50})/10,000]^{1/2}$, etc. The results show that all the coverage probabilities of $\bar{P}_\theta(T^\dagger_{n2} \geq c_n)$, $\bar{P}_\theta(T^\dagger_{n2} \leq -c_n)$ and $\bar{P}_\theta(|T^\dagger_{n2}| \leq c_n)$ are within 1.96 standard deviations, where $\bar{P}_\theta$ denotes the joint probability distribution of the data and the simulated sample. Moreover, all of the coverage probabilities for (39) are within 0.001 of those for (20).

4.2. A TAR(1) example

Consider the TAR model in (3). For $\{y_t\}$ to be ergodic, the parameter space is determined by the inequalities: $\theta_1 < 1, \theta_2 < 1$, and $\theta_1\theta_2 < 1$; see, for example, Petruccelli and Woolford (1984) and Chen and Tsay (1991). Recall the Q_θ matrix in (28): $q_{11} = (E_\theta(S^+)^2)^{-1/2}$, $q_{12} = q_{21} = 0$ and $q_{22} = (E_\theta(S^-)^2)^{-1/2}$. So, $\delta_{ij}(\theta)$ and $\mu_n(\theta)$ in (11) and (13) have simpler forms: $\mu_n(\theta) = (-q^{\#}_{11}(\theta)/\sqrt{n}, -q^{\#}_{22}(\theta)/\sqrt{n})'$ and $\delta_{ij} = q_{ii,j}(\theta)q_{jj,i}(\theta)$.

Below we briefly describe the procedures of bootstrap methods. For parametric bootstrap, we form a bootstrap sample $\{y_1^{(b)}, \dots, y_n^{(b)}\}$ by

$$y_t^{(b)} = \hat{\theta}_{n1}(y_{t-1}^{(b)})^+ + \hat{\theta}_{n2}(y_{t-1}^{(b)})^- + \varepsilon_t^{(b)}, \tag{41}$$

where $y_0^{(b)} = 0, t = 1, \dots, n$, and $\{\varepsilon_1^{(b)}, \dots, \varepsilon_n^{(b)}\}$ is a random sample from standard normal distribution. Then, the bootstrap percentile and bootstrap-t intervals are constructed based on B bootstrap samples. That is, the approximate level γ confidence interval for $\theta_i, i = 1, 2$ by the bootstrap-t is

$$\bar{\theta}_i \pm s_i c_n,$$

where

$$\bar{\theta}_i = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_i^{(b)}, \quad s_i = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}_i^{(b)} - \bar{\theta}_i)^2},$$

and c_n is the $100(1+\gamma)/2$ quantile of the standard univariate t-distribution with n degrees of freedom, and that by the bootstrap percentile is

$$(\hat{\theta}_{(i,l(\gamma))}, \hat{\theta}_{(i,u(\gamma))}),$$

where $\hat{\theta}_{(i,j)}$ is the j th order statistic of $\{\hat{\theta}_i^{(1)}, \dots, \hat{\theta}_i^{(B)}\}$, $l(\gamma) = \lfloor B \cdot (1-\gamma)/2 \rfloor$ and $u(\gamma) = \lfloor B \cdot (1+\gamma)/2 \rfloor$, and $\lfloor a \rfloor$ denotes the largest integer less than or equal to a. The procedures of non-parametric bootstrap are the same as the parametric one except that the $\varepsilon_t^{(b)}$ in (41) are obtained by bootstrapping the residuals. In the experiment we set $B = 1000$.

In related work, Enders et al. (2007) compared the coverage probabilities in TAR models by the normal approximation and bootstrap methods. They reported the actual coverage attained and a measure of the symmetry of the intervals. The symmetry measure was constructed as follows. For the ideal 90% confidence interval, we expect the true parameter to fall below the lower bound of the interval five percent of the time and above the upper bound of the interval five percent of the time. Let A_1 denote the actual percent of time that the parameter falls above the upper bound and A_2 denote the actual percent of time it falls below the lower bound. Then the symmetry is set as the value $[(A_1-5)^2 + (A_2-5)^2]^{1/2}$. For the 95% confidence interval, the symmetry measure can be defined similarly.

In simulating TAR processes, it is possible that the constructed series never crosses the true threshold, especially when the slope parameter is close to 1. In such cases, the $X_n'X_n$ is not positive definite and it is impossible to fit a TAR model. In Enders et al. (2007), if a simulated series did not contain at least three points on each side of the threshold, it was discarded and replaced with another simulated sample. They applied this rule also for bootstrap simulation. Instead of removing such series, an alternative approach is to regularize the information matrix $X_n'X_n$ by adding a small positive quantity in the diagonal elements of the matrix, which amounts to put a tiny proportion of observations to both regimes. The idea of regularizing the information matrix is not new; see, for example, Vuchkov (1977). In unreported experiments we found that the results of the two approaches are similar; and since the regularizing approach makes more efficient use of the sample, we take this approach (and set the small positive quantity as 0.001) in the experiments below.

We conduct simulation study with $(\theta_1, \theta_2) = (0.3, 0.3), (0.3, 0.6), (0.3, 0.9), (0.3, 0.95), (0.6, 0.6), (0.6, 0.9), (0.6, 0.95), (0.9, 0.9), (0.9, 0.95), (0.95, 0.95)$. These parameter values are the same as in Table 2 of Enders et al. (2007). As $T_n^\dagger$ depends on the choice of (η, m), we conducted simulation with $(\eta, m) = (0.001, 1000)$ and $(0.0005, 5000)$ and found that $T_n^\dagger$ is not sensitive to these values. So, for the rest of the experiments we take $(\eta, m) = (0.001, 1000)$. To compare with Table 2 of Enders et al. (2007), for each selected parameter value, we generate 10,000 realizations of $y_1, \dots, y_n$ for $n = 100$. The initial value y_0 was set to be 0 and ε_t's were drawn from the standard normal distribution. For each realization, the model was fitted using maximum likelihood estimate (MLE), equivalent to least squares estimate in this case. The estimates were used to construct the 90% confidence intervals for θ_1 and θ_2 using the uncorrected pivot T_n in (6) and the corrected pivot $T_n^\dagger$ (27), and to generate 1000 bootstrap samples to form the bootstrap percentile and bootstrap-t intervals.

Table 2a gives simulated coverage probabilities and the symmetry measures (in parentheses). In this table, the terms boot-t and boot-p represent bootstrap-t and bootstrap percentile, respectively; and the (n) stands for nonparametric and (p) for parametric. We made several observations. First, it is not difficult to check that the normal approximation in Enders et al. (2007) (based on

Table 2a
TAR(1) model ($n=100$, replicates $=10,000$) (nominal Coverage $=90\%$; $c_n=1.6602$).

$(\theta_1\ \theta_2)$	Uncorrected		Corrected		Boot-$t(n)$		Boot-$t(p)$		Boot-$p(n)$		Boot-$p(p)$	
(0.30 0.30)	0.903	0.899	0.899	0.897	0.889	0.881	0.908	0.908	0.872	0.870	0.892	0.893
	(1.41)	(0.93)	(0.18)	(0.28)	(3.38)	(3.50)	(3.83)	(3.55)	(5.80)	(5.81)	(5.88)	(5.62)
(0.30 0.60)	0.901	0.899	0.899	0.896	0.896	0.883	0.908	0.906	0.871	0.852	0.885	0.875
	(1.85)	(1.57)	(0.41)	(0.29)	(4.17)	(4.84)	(4.46)	(5.25)	(7.14)	(8.60)	(7.00)	(8.31)
(0.30 0.90)	0.900	0.893	0.900	0.896	0.939	0.887	0.927	0.899	0.851	0.790	0.859	0.796
	(2.36)	(3.28)	(0.45)	(0.38)	(4.33)	(6.66)	(5.12)	(7.08)	(10.4)	(16.1)	(10.1)	(16.2)
(0.30 0.95)	0.904	0.890	0.905	0.898	0.965	0.890	0.948	0.894	0.834	0.754	0.835	0.750
	(2.59)	(4.16)	(0.43)	(0.54)	(4.97)	(7.05)	(4.91)	(7.30)	(12.3)	(19.8)	(12.4)	(20.5)
(0.60 0.60)	0.897	0.900	0.897	0.899	0.891	0.888	0.907	0.908	0.844	0.845	0.862	0.863
	(2.55)	(1.84)	(0.46)	(0.09)	(5.25)	(5.42)	(5.90)	(5.62)	(10.3)	(10.1)	(9.78)	(9.64)
(0.60 0.90)	0.898	0.891	0.900	0.898	0.938	0.894	0.927	0.901	0.813	0.772	0.822	0.778
	(3.07)	(3.69)	(0.59)	(0.43)	(4.56)	(6.65)	(5.43)	(6.98)	(14.3)	(18.0)	(13.7)	(17.9)
(0.60 0.95)	0.899	0.887	0.902	0.900	0.963	0.894	0.948	0.899	0.783	0.733	0.787	0.730
	(3.14)	(4.43)	(0.43)	(0.61)	(5.02)	(6.81)	(4.96)	(7.00)	(17.2)	(21.9)	(17.0)	(22.4)
(0.90 0.90)	0.888	0.887	0.901	0.903	0.942	0.939	0.929	0.928	0.712	0.707	0.717	0.716
	(4.70)	(4.50)	(0.99)	(0.72)	(4.80)	(4.67)	(5.42)	(5.46)	(24.1)	(24.5)	(23.8)	(23.9)
(0.90 0.95)	0.887	0.882	0.904	0.905	0.966	0.941	0.952	0.930	0.661	0.659	0.670	0.659
	(4.90)	(5.30)	(1.07)	(1.16)	(5.08)	(4.63)	(5.00)	(5.28)	(29.1)	(29.2)	(28.4)	(29.4)
(0.950.95)	0.880	0.877	0.906	0.909	0.964	0.966	0.953	0.949	0.606	0.609	0.612	0.614
	(5.97)	(6.03)	(1.19)	(1.44)	(4.95)	(5.01)	(4.96)	(4.91)	(34.5)	(34.2)	(34.1)	(33.8)

Table 2b
TAR(1) model ($n=50$, replicates $=10,000$) (nominal Coverage $=95\%$; $c_n=2.0086$; $\pm$ is the range within 1.96 standard deviations).

$(\eta,m)=(0.001,1000)$ $(\theta_1\ \theta_2)$	Upper	Lower	CI	Symm	Upper	Lower	CI	Symm
	T_{n1}	T_{n1}	T_{n1}	T_{n1}	T_{n2}	T_{n2}	T_{n2}	T_{n2}
	$T_{n1}^{\dagger}$	$T_{n1}^{\dagger}$	$T_{n1}^{\dagger}$	$T_{n1}^{\dagger}$	$T_{n2}^{\dagger}$	$T_{n2}^{\dagger}$	$T_{n2}^{\dagger}$	$T_{n2}^{\dagger}$
(0.30 0.30)	0.028	0.015	0.956	1.02	0.031	0.020	0.950	0.79
	0.025	0.023	0.952	0.20	0.026	0.027	0.947	0.20
(0.30 0.60)	0.030	0.015	0.954	1.12	0.035	0.016	0.949	1.35
	0.026	0.022	0.952	0.30	0.027	0.026	0.947	0.22
(0.30 0.90)	0.037	0.012	0.951	1.80	0.044	0.011	0.945	2.41
	0.027	0.021	0.951	0.45	0.029	0.018	0.954	0.80
(0.30 0.95)	0.037	0.011	0.952	1.87	0.047	0.010	0.943	2.71
	0.026	0.022	0.952	0.31	0.029	0.015	0.957	1.06
(0.60 0.60)	0.033	0.012	0.954	1.52	0.036	0.015	0.949	1.46
	0.025	0.021	0.953	0.37	0.027	0.024	0.949	0.23
(0.60 0.90)	0.040	0.010	0.950	2.09	0.042	0.011	0.947	2.25
	0.027	0.019	0.954	0.62	0.029	0.016	0.955	0.96
(0.60 0.95)	0.042	0.009	0.949	2.33	0.047	0.009	0.943	2.72
	0.027	0.021	0.953	0.48	0.030	0.014	0.956	1.16
(0.90 0.90)	0.047	0.008	0.945	2.78	0.043	0.009	0.948	2.40
	0.027	0.015	0.957	1.00	0.026	0.015	0.959	1.04
(0.90 0.95)	0.048	0.007	0.945	2.87	0.048	0.009	0.944	2.79
	0.028	0.017	0.955	0.85	0.026	0.014	0.960	1.09
(0.95 0.95)	0.051	0.008	0.941	3.11	0.046	0.008	0.946	2.68
	0.029	0.019	0.952	0.72	0.023	0.013	0.964	1.25
$\pm$	0.025 ± 0.003	0.025 ± 0.003	0.95 ± 0.004		0.025 ± 0.003	0.025 ± 0.003	0.95 ± 0.004	

least squares t-statistic) is exactly our uncorrected pivot; and indeed our experimental results for the uncorrected pivot agree with theirs in both the coverage probabilities and the measure of symmetry. For the corrected pivot $T_n^{\dagger}$, we note that when the true parameter is near 1 and n is small, it is quite often to get $\hat{\theta}_{ni}>1$, which is not in the assumed parameter space; therefore, the correction terms $\tilde{\mu}_n$ and $\tilde{\Delta}_\theta$ in (26) (to be obtained by simulated samples with underlying parameter $\hat{\theta}_n$ and $\hat{\theta}_n+\eta e_k$) are not

Table 2c
TAR(1) model (replicates = 10, 000).

$(\eta, m) = (0.001, 1000)$	$n=50$			$n=100$		
$(\theta_1\ \theta_2)$	$\#\{n_1=0 \text{ or } n_2=0\}$	$\#\{\hat{\theta}_{n1}>1\}$	$\#\{\hat{\theta}_{n2}>1\}$	$\#\{n_1=0 \text{ or } n_2=0\}$	$\#\{\hat{\theta}_{n1}>1\}$	$\#\{\hat{\theta}_{n2}>1\}$
(0.30 0.30)	0	0	0	0	0	0
(0.30 0.60)	0	0	0	0	0	0
(0.30 0.90)	36	5	66	0	0	1
(0.30 0.95)	211	11	382	25	1	47
(0.60 0.60)	0	1	1	0	0	0
(0.60 0.90)	36	14	89	0	1	3
(0.60 0.95)	211	19	420	25	2	57
(0.90 0.90)	80	157	176	0	15	15
(0.90 0.95)	255	166	503	25	31	110
(0.95 0.95)	418	464	490	48	123	148

available. There are no standard ways to handle these samples. We simply set $\hat{\theta}_{ni} = 0.95$ and then run Monte Carlo simulation to get the correction terms. The obtained results show that the corrected pivot is more symmetric and the overall coverage probabilities are closer to the nominal value than the uncorrected one. For bootstrap methods, we observe that overall the coverage probabilities by bootstrap-t are closer to the nominal value than that by bootstrap percentile, and that the coverage probabilities by bootstrap percentile are far below the nominal value when the parameters are closer to 1. These agree with the findings in Enders et al. (2007). However, our bootstrap-t are more conservative than theirs when the parameter is close to 1. This may be due to different treatments for cases such as $\hat{\theta}_{ni} > 1$ or unbalanced numbers of observations in both regimes. We also conduct simulation for T_{ni} and $T^{\dagger}_{ni}$ with $n = 50$ and nominal level 0.95. The results are in Table 2b, where the term symm denotes the symmetry measure, and the terms upper, lower, and CI represent simulated upper 0.025, lower 0.025, and the 95% coverage probabilities, respectively. This table shows that T_{ni} has much larger upper tail than lower tail, indicating that $\hat{\theta}_{ni}$ tends to be biased downwards at current θ values. In contrast, the two tails of $T^{\dagger}_{ni}$ are more balanced. Note that here c_n denotes the 2.5th or 5th percentile of the standard univariate t-distribution with n degrees of freedom.

Finally, we provide in Table 2c the number of samples where one of the regimes contains no observations, and the number of samples where $\hat{\theta}_{ni} > 1$. In this table, n_1 and n_2 denote the numbers of observations in regimes 1 and 2, and the sample sizes are set as $n = 50$ and 100. Obviously, the proportions of such samples are higher when the sample size is smaller and the parameter values are close to 1.

5. Conclusion

We have studied the very weak type approximation to obtain corrected confidence sets for the parameters by approximating the mean and variance corrections numerically; and we have shown that the confidence sets are accurate to $o_p(1/n)$ in the sense of (31). This numerical approximation involves two tuning quantities η and m. Our experiments showed that these two quantities are easy to tune and the experiment results are not sensitive to the choice of them. The proposed approach was applied to AR(2) and TAR(1) models and the results are mostly satisfactory, except that special concerns are needed when the parameter values are close to 1. In the future, it would be of interest to extend this approach to general TAR models and some adaptive nonlinear models.

Acknowledgments

The authors would like to thank the referees for their valuable comments on the paper. The authors are partially supported by the National Science Council of Taiwan. The last two authors are also partially supported by Center for Service Innovation at National Chengchi University.

References

Anděl, J., Bartoň, T., 1986. A note on the threshold AR(1) model with Cauchy innovations. J. Time Ser. Anal. 7, 1–5.
Anděl, J., Netuka, I., Zvára, K., 1984. On threshold autoregressive processes. Kybernetika 20, 89–106.
Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods. Springer, New York.
Chen, B.R., Tsay, R.S., 1991. On the ergodicity of TAR(1) processes. Ann. Appl. Probab. 1, 613–634.
Coad, D.S., Woodroofe, M., 1998. Approximate bias calculations for sequentially designed experiments. Sequential Anal. 17, 1–31.
Eisele, J.R., 1994. The doubly adaptive biased coin design for sequential clinical trials. J. Statist. Plann. Inference 38, 249–262.
Enders, W., Falk, B., Siklos, P.L., 2007. A threshold model of real U.S. GDP and the problem of constructing confidence intervals in TAR models. Stud. Nonlinear Dynamics Econometrics 11 (3), 1322.
Lai, T.L., Wei, C.Z., 1982. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. Ann. Statist. 10, 154–166.
Loges, W., 2004. The stationary marginal distribution of a threshold AR(1) process. J. Time Ser. Anal. 25, 103–125.
Petruccelli, J.D., Woolford, S.W., 1984. A threshold AR(1) model. J. Appl. Probab. 21, 270–286.
Stein, C., 1981. Estimation of the mean of a multivariate normal distribution. Ann. Statist. 9, 1135–1151.

Tong, H., 1983. Threshold Models in Non-linear Time Series. Springer, New York.
Tong, H., 1990. Non-linear Time Series: A Dynamical System Approach. Oxford University Press, Oxford.
Vuchkov, I.N., 1977. A ridge-type procedure for design of experiments. Biometrika 64, 147–150.
Weng, R.C., Woodroofe, M., 2006. Approximate confidence sets for a stationary AR(p) process. J. Statist. Plann. Inference 136, 2719–2745.
Woodroofe, M., 1986. Very weak expansions for sequential confidence levels. Ann. Statist. 14, 1049–1067.
Woodroofe, M., 1989. Very weak expansions for sequentially designed experiments: linear models. Ann. Statist. 17, 1087–1102.
Woodroofe, M., Coad, D.S., 1997. Corrected confidence sets for sequentially designed experiments. Statist. Sinica 7, 53–74.
Woodroofe, M., Coad, D.S., 1999. Corrected confidence sets for sequentially designed experiments: examples. In: Ghosh, S. (Ed.), Multivariate Analysis, Design of Experiments, and Survey Sampling. Marcel Dekker, Inc., New York, pp. 135–161.