Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Statistical Planning and Inference 136 (2006) 2719–2745
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Approximate confidence sets for a stationary AR(p) process

Ruby C. Weng[a,*], Michael Woodroofe[b]

[a]*National Chengchi University, Taipei 11623, Taiwan*
[b]*University of Michigan, Ann Arbor, MI 48109-1027, USA*

Received 14 May 2004; received in revised form 14 November 2004; accepted 14 November 2004
Available online 9 January 2005

Abstract

Approximate confidence intervals are derived for the autoregressive parameters of a stationary, Gaussian auto-regressive process of arbitrary order and shown to be asymptotically correct to order $o(1/n)$, where n is the sample size. Simulation studies are included for small and moderate sample sizes for the case of two auto-regressive parameters, and these indicate excellent approximation for sample sizes as small as $n = 10, 20$. The convergence is in the very weak sense, and the derivation differs from most existing work through its direct focus on Studentized estimation error and its use of Stein's identity.

Keywords: Asymptotic expansions; Asymptotic confidence levels; Stationary autoregressive process; Very weak expansions

1. Introduction

In 1943 Mann and Wald wrote two pioneering papers. In Mann and Wald (1943a) they introduced stochastic order relations which have since become an integral part of large sample theory. In Mann and Wald (1943b) they gave a careful proof of the consistency and asymptotic normality of the maximum likelihood estimator from a stationary autoregressive process, a topic which is now an integral part of advanced courses on time series—for example, Brockwell and Davis (1991, pp. 258–262). Here we continue the line begun in

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

0378-3758/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2004.11.007

Mann and Wald (1943b) by obtaining asymptotic expansions for the distribution of the maximum likelihood estimator suitably normalized.

Consider a stationary autoregressive process of order p,

$$y_t = \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + \sigma e_t, \quad t = 0, \pm 1, \ldots,$$

where e_t are independent standard normal random variables and $\theta_1, \ldots, \theta_p \in \Re$ and $\sigma^2 > 0$ are unknown parameters. We suppose throughout that the process is causal. Thus, let Ω be all $\omega = (\omega_1, \ldots, \omega_p)'$ for which the polynomial $1 - (\omega_1 z + \cdots + \omega_p z^p)$ does not vanish for complex $|z| \leqslant 1$, and suppose that $\theta = (\theta_1, \ldots, \theta_p)' \in \Omega$. It is well known that the maximum likelihood estimators may be severely biased in the AR(p) models. See, for example, Coad and Woodroofe (1998). The purpose of this paper is to derive an asymptotic expansion for the distribution of the maximum likelihood estimator, suitably renormalized, up to terms that are small compared to $1/n$, where n is the sample size. Our findings are similar to those of Tanaka (1984), who derived Edgeworth expansions for ARMA models, but there are several important differences. The normalization used here is different, employing a random matrix instead of $\sqrt{n}$. See (3), below. This normalization leads to a simple expansion. It is possible to describe the expansion solely in terms of the mean and covariance matrix of the renormalized estimation error, and this simplifies the formation of confidence sets. Further, the derivation here follows work in Woodroofe and Coad (1997, 2002) on sequentially designed experiments and is entirely different from Tanaka (1984). On the other hand, Tanaka's model is more general.

The nature of the expansions and their derivations is presented in Sections 2 and 3, and the expansions are illustrated by an example in Section 4. Sections 5 and 6 contain details of the proofs. Some discussions are made in Section 7.

2. Preliminaries

The Likelihood Function. To fix ideas, suppose that $y_1, \ldots, y_{p+n}$ are observed, and let

$$X_n = \begin{pmatrix} y_p & \cdot & \cdot & y_1 \\ y_{p+1} & \cdot & \cdot & y_2 \\ \cdot & \cdot & \cdot & \cdot \\ y_{p+n-1} & \cdot & \cdot & y_n \end{pmatrix}$$

$e_{k,n} = [e_{k+1}, \ldots, e_{k+n}]'$, and $y_{k,n} = [y_{k+1}, \ldots, y_{k+n}]'$, where $'$ denotes transpose. Then the model may be written as

$$y_{p,n} = X_n \theta + \sigma e_{p,n},$$

$$y_{0,p} \sim N_p(0, \sigma^2 G_\theta),$$

the normal distribution with mean 0 and covariance matrix $\sigma^2 G_\theta = E_{\sigma,\theta}(y_{0,p} y'_{0,p})$. So, the log-likelihood function given $(y_1, \ldots, y_{p+n})'$ is

$$\ell_n(\sigma^2, \theta) = \ell_0(\sigma^2, \theta) - \frac{1}{2\sigma^2} \|y_{p,n} - X_n \theta\|^2 - \frac{1}{2} n \log(\sigma^2),$$

where

$$\ell_0(\sigma^2,\theta) = -\frac{1}{2}\log(\det G_\theta) - \frac{1}{2\sigma^2} y'_{0,p} G_\theta^{-1} y_{0,p} - \frac{1}{2} p \log(\sigma^2)$$

depends only on the first p observations. Further,

$$\nabla \ell_n(\sigma^2,\theta) = \frac{1}{\sigma^2} X'_n (y_{p,n} - X_n\theta) + \nabla \ell_0(\sigma^2,\theta),$$
$$\nabla^2 \ell_n(\sigma^2,\theta) = -\frac{1}{\sigma^2} X'_n X_n + \nabla^2 \ell_0(\sigma^2,\theta),$$

where ∇ denotes differentiation with respect to θ. It is easily seen (and follows from Lemma 5.2 below) that the maximum likelihood estimators, $\hat{\theta}_n$ and $\hat{\sigma}_n^2$ say, exist *w.p.*1 and satisfy the likelihood equation. So,

$$\hat{\theta}_n = (X'_n X_n)^{-1}[X'_n y_{p,n} + \hat{\sigma}_n^2 \nabla \ell_0(\hat{\sigma}_n^2, \hat{\theta}_n)]$$

and

$$\hat{\sigma}_n^2 = \frac{\|y_{p,n} - X_n\hat{\theta}_n\|^2 + y'_{0,p} G_{\hat{\theta}_n}^{-1} y_{0,p}}{n+p}. \tag{1}$$

Writing $\ell_n(\sigma^2,\theta) = \ell_n(\sigma^2,\theta|y_{0,n+p})$ to emphasize the dependence on $y_1,\ldots,y_{n+p}$, it is easily seen that $\ell_n(\sigma^2,\theta|cy_{0,n+p}) = \ell_n(c^{-2}\sigma^2,\theta|y_{0,n+p}) - (n+p)\log(c)$ for all $c>0$. So, $\hat{\theta}_n$ is invariant under scale transformations and $\hat{\sigma}_n$ is equivariant, i.e. $\hat{\theta}_n(cy_{0,n+p}) = \hat{\theta}_n(y_{0,n+p})$ and $\hat{\sigma}_n^2(cy_{0,n+p}) = c^2\hat{\sigma}_n^2(y_{0,n+p})$ for all $c>0$. The least squares estimators are denoted by $\tilde{\theta}_n$ and $\tilde{\sigma}_n^2$, so that

$$\tilde{\theta}_n = (X'_n X_n)^{-1} X'_n y_{p,n}$$

$$\tilde{\sigma}_n^2 = \frac{\|y_{p,n} - X_n\tilde{\theta}_n\|^2}{n-p}$$

for $n>p$. These are similarly invariant and equivariant.

Very weak expansions: Let $\hat{\mathscr{I}}_n$ denote the information matrix, $\hat{\mathscr{I}}_n = -\nabla^2\ell_n(\hat{\sigma}_n^2,\hat{\theta}_n)$. Then $\hat{\mathscr{I}}_n$ is invariant too, and $\hat{\mathscr{I}}_n$ is non-negative definite $w.p.1$. Let $B_n = B_n(y_{0,n+p})$ be a scale equivariant $p\times p$ matrix for which

$$B_n B'_n = \hat{\sigma}_n^2 \hat{\mathscr{I}}_n = X'_n X_n - \hat{\sigma}_n^2 \nabla^2 \ell_0(\hat{\sigma}_n^2,\hat{\theta}_n). \tag{2}$$

There are many possible choices for B_n. The main requirements are (2) and (12) below. Let T_n be the studentized estimation error

$$T_n = \frac{1}{\tilde{\sigma}_n} B'_n(\theta - \hat{\theta}_n). \tag{3}$$

A main result provides an asymptotic expansion for the distribution of T_n from which improved confidence intervals may be found. The derivation proceeds by first considering the related quantity

$$Z_n = \frac{1}{\sigma} B_n'(\theta - \hat{\theta}_n). \tag{4}$$

The distributions of T_n and Z_n do not depend on σ. *So, there is no loss of generality in supposing that* $\sigma = 1$ *when studying them; and* σ *is omitted from the notation in the sequel, so that* E_θ *is written for* $E_{1,\theta}$, $\ell_n(\theta)$ *for* $\ell_n(1, \theta)$, *etc.* It is shown that to order $o(1/n)$, Z_n is normal with a mean $\mu_n(\theta)$ and a covariance matrix $\Sigma_n(\theta)$ that are approaching 0 and the identity matrix. Further, there are estimators $\hat{\mu}_n$ and $\hat{\Gamma}_n$ for which $Z_n^* := \hat{\Gamma}_n^{-1}(Z_n - \hat{\mu}_n)$ is asymptotically standard normal to order $o(1/n)$, and the distribution of $T_n^* := \hat{\Gamma}_n^{-1}(T_n - \hat{\mu}_n)$ differs from a p-variate t-distribution with n degrees of freedom by $o(1/n)$.

The convergence here is in the very weak sense of Woodroofe (1986, 1989). In the case of Z_n^*, this means that

$$\int_\Omega [P_\theta\{Z_n^* \in B\} - \Phi^p(B)]\xi(\theta)\,d\theta = o\left(\frac{1}{n}\right) \tag{5}$$

uniformly with respect to Borel sets $B \subseteq \Re^p$ for all twice continuously differentiable densities ξ with compact convex support in Ω that satisfy the mild condition (31) below, where Φ^p is the standard p-variate normal distribution. Woodroofe (1989) writes (5) as

$$P_\theta\{Z_n^* \in B\} = \Phi^p(B) + o\left(\frac{1}{n}\right) \tag{6}$$

very weakly, and argues that (6) is strong enough to support a frequentist interpretation for confidence intervals. The corresponding result for T_n is

$$P_\theta\{T_n^* \in B\} = G_n^p(B) + o\left(\frac{1}{n}\right) \tag{7}$$

very weakly, uniformly with respect to Borel sets $B \subseteq \Re^p$, where G_n^p is the spherically symmetric p-variate t distribution with n degrees of freedom. It is easy to use (7) to form corrected confidence sets. The procedure is illustrated in Section 4.

The Bayesian connection: The integrated probability in (5) is probability in a Bayesian model in which θ is given a prior density ξ. So, consider a Bayesian model in which θ has a twice continuously differentiable prior density ξ with compact convex support $K \subseteq \Omega$. Throughout this paper we denote the probability and expectation corresponding to prior ξ as P_ξ and E_ξ, and the conditional expectation given $y_1, \ldots, y_{p+n}$ as E_ξ^n. When $\sigma = 1$, the posterior density of θ given $y_1, \ldots, y_{p+n}$ is

$$\xi_n(\theta) \propto e^{\ell_n(\theta) - \ell_n(\hat{\theta}_n)} \xi(\theta).$$

From (2) and (4)

$$\ell_n(\theta) - \ell_n(\hat{\theta}_n) = -\tfrac{1}{2}\|Z_n\|^2 + r_n(\theta),$$

where

$$r_n(\theta) = \ell_n(\theta) - \ell_n(\hat{\theta}_n) - \tfrac{1}{2}(\theta - \hat{\theta}_n)'\hat{\sigma}_n^2\nabla^2\ell_n(\hat{\sigma}_n^2, \hat{\theta}_n)(\theta - \hat{\theta}_n). \tag{8}$$

So, the posterior density of Z_n is

$$\zeta_n(z) \propto \phi_p(z) f_n(z), \tag{9}$$

where

$$f_n(z) = \xi(\theta)\mathrm{e}^{r_n(\theta)},$$

θ and z are related by (4), and ϕ_p is the standard p-variate normal density. For later reference, observe that

$$\frac{\nabla f_n}{f_n}(z) = B_n^{-1}\left[\frac{\nabla\xi}{\xi}(\theta) + \nabla r_n(\theta)\right] \tag{10}$$

and

$$\frac{\nabla^2 f_n}{f_n} = B_n^{-1}\left[\frac{\nabla^2\xi}{\xi} + \frac{\nabla\xi}{\xi}\nabla r_n' + \nabla r_n\frac{\nabla\xi'}{\xi} + \nabla^2 r_n + \nabla r_n\nabla r_n'\right]B_n'^{-1}, \tag{11}$$

where $\nabla f_n(z)$ is obtained by differentiation with respect to z, and $\nabla\xi(\theta)$ and $\nabla r_n(\theta)$ by differentiation with respect to θ.

Stein's identity: The basic approach makes use of Stein's (1981, 1987) identity, which is reviewed next. Recall that Φ^p denotes the standard p-variate normal distribution and write

$$\Phi^p h = \int h\,\mathrm{d}\Phi^p$$

for functions h for which the integral is finite. Next let Γ denote a finite signed measure of the form $\mathrm{d}\Gamma = f\,\mathrm{d}\Phi^p$, where f is a real-valued function defined on $\Re^p$ satisfying $\Phi^p|f| = \int|f|\,\mathrm{d}\Phi^p < \infty$. The posterior density of Z_n is of this form by (9). For $s > 0$, let $\mathscr{H}_s^o$ be the collection of all measurable functions $h : \Re^p \to \Re$ for which $|h(z)| \leqslant 1 + \|z\|^s$; let $\mathscr{H}_s = \{h : h/b \in \mathscr{H}_s^o,$ for some $b > 0\}$; and let $\mathscr{H} = \bigcup_{s\geqslant 0}\mathscr{H}_s$. Given $h \in \mathscr{H}$, let $h_0 = \Phi^p h$, $h_p = h$,

$$h_j(y_1, \ldots, y_j) = \int_{\Re^{p-j}} h(y_1, \ldots, y_j, w)\Phi_{p-j}(\mathrm{d}w),$$

for $j = 1, \ldots, p-1$, and

$$\begin{aligned} &g_j(y_1, \ldots, y_p) \\ &\quad = \mathrm{e}^{1/2y_j^2}\int_{y_j}^{\infty}[h_j(y_1, \ldots, y_{j-1}, w) - h_{j-1}(y_1, \ldots, y_{j-1})]\mathrm{e}^{-1/2w^2}\,\mathrm{d}w \end{aligned}$$

for $-\infty < y_1, \ldots, y_p < \infty$ and $j = 1, \ldots, p$. Each g_j is regarded as a function on $\Re^p$, even though g_j only depends on $y_1, \ldots, y_j$. Next, let

$$Uh = (g_1, \ldots, g_p)'.$$

The transformation U may be iterated. Let U^2h be the $p \times p$ matrix whose jth column is Ug_j, and let

$$Vh = \frac{(U^2h + U^2h')}{2}.$$

Then *Vh* is a symmetric matrix. Simple calculations show that

$$\Phi^p(Uh) = \int_{\Re^p} zh(z)\Phi^p(\mathrm{d}z)$$

and

$$\Phi^p(Vh) = \frac{1}{2}\int_{\Re^p} (zz' - I_p)h(z)\Phi^p(\mathrm{d}z)$$

for all $h \in \mathscr{H}$. When $p = 1$, these formulas simplify. Then

$$Uh(z) = \mathrm{e}^{1/2z^2} \int_z^\infty (h(y) - \Phi h)\mathrm{e}^{-1/2w^2}\,\mathrm{d}w$$

and U^2 is the composition of U with itself. It may be shown that if $h \in \mathscr{H}_s$, then $\|Uh\| \in \mathscr{H}_{s'}$, where $s' = \max(0, s-1)$. See Woodroofe (1992).

Proposition 2.1 (*Stein's identity*). *Let r be a nonnegative integer. Suppose that* $\mathrm{d}\Gamma = f\,\mathrm{d}\Phi^p$, *where f is a differentiable function on* $\Re^p$, *for which*

$$\int_{\Re^p} |f|\,\mathrm{d}\Phi^p + \int_{\Re^p} (1 + \|z\|^r)\|\nabla f(z)\|\Phi^p(\mathrm{d}z) < \infty,$$

then

$$\Gamma h = \Gamma 1 \cdot \Phi^p h + \int_{\Re^p} Uh(z)'\nabla f(z)\Phi^p(\mathrm{d}z)$$

for all $h \in \mathscr{H}_r$. *If* $\partial f/\partial z_j,\ \ j = 1, \ldots, p$, *are differentiable, and*

$$\int_{\Re^p} (1 + \|z\|^r)\|\nabla^2 f(z)\|\Phi^p(\mathrm{d}z) < \infty$$

then, for all $h \in \mathscr{H}_r$,

$$\Gamma h = \Gamma 1 \cdot \Phi^p h + \Phi^p(Uh)' \int_{\Re^p} \nabla f(z)\Phi^p(\mathrm{d}z) + \int_{\Re^p} \mathrm{tr}[(Vh)\nabla^2 f]\,\mathrm{d}\Phi^p.$$

Proof. See Woodroofe (1989, Proposition 1) and Woodroofe and Coad (1997, Proposition 2). □

3. Main results

In this section we state the results of the paper and outline the proofs. The details of the proofs are deferred to Section 6. In addition to (2), it is required that B_n be so chosen that

$$Q_n := \sqrt{n} B_n^{-1} \to Q_\theta \tag{12}$$

in P_θ-probability for all θ when $\sigma = 1$, where the entries in Q_θ are twice continuously differentiable in θ. This will always be true if $B_n B_n'$ is a Cholesky decomposition of $\hat{\sigma}_n^2 \mathscr{I}_n$. For then

$$\frac{B_n B_n'}{n} \to G_\theta$$

$w.p.1(P_\theta)$, where G_θ is the covariance matrix of $y_1, \ldots, y_p$. Writing $G_\theta^{-1} = Q_\theta' Q_\theta$ by a Cholesky decomposition, the entries of Q_θ are twice continuously differentiable and $\lim_{n\to\infty} Q_n = Q_\theta$, $w.p.1.(P_\theta)$, for all $\theta \in \Omega$.

In Section 5, events D_n are constructed for which $P_\xi(D_n^c) = \mathrm{o}(1/n)$ and the likelihood function is well behaved when D_n occur. So, if h is a bounded measurable function, then

$$E_\xi h(Z_n) = E_\xi\{E_\xi^n[h(Z_n)]1_{D_n}\} + \mathrm{o}\left(\frac{1}{n}\right).$$

Restricting attention to D_n and applying Stein's identity to the posterior distribution of $h(Z_n)$ given $y_1, \ldots, y_{p+n}$,

$$E_\xi^n\{h(Z_n)\} = \Phi^p h + (\Phi^p U h)' E_\xi^n \left\{\frac{\nabla f_n}{f_n}(Z_n)\right\} + E_\xi^n \left\{\mathrm{tr}\left[V h(Z_n) \frac{\nabla^2 f_n}{f_n}(Z_n)\right]\right\}.$$

Using (10) and (11), this may be written

$$\begin{aligned} E_\xi^n\{h(Z_n)\} = \Phi^p h &+ \frac{1}{\sqrt{n}} (\Phi^p U h)' E_\xi^n \left\{Q_\theta \left(\frac{\nabla \xi}{\xi}\right)\right\} \\ &+ \frac{1}{n} E_\xi^n \left\{\mathrm{tr}\left[(\Phi^p V h) Q_\theta \left(\frac{\nabla^2 \xi}{\xi}\right) Q_\theta'\right]\right\} \\ &+ \frac{1}{\sqrt{n}} (\Phi^p U h)' [\mathrm{I}_n + \mathrm{II}_n] + \frac{1}{n} [\mathrm{III}_n(h) + \mathrm{IV}_n(h)], \end{aligned} \tag{13}$$

where

$$\mathrm{I}_n = E_\xi^n \left\{(Q_n - Q_\theta) \left(\frac{\nabla \xi}{\xi}\right)\right\},$$

$$\mathrm{II}_n = E_\xi^n\{Q_n \nabla r_n\},$$

$$\mathrm{III}_n(h) = E_\xi^n \left\{\mathrm{tr}\left[V h(Z_n) Q_n \left(\frac{\nabla \xi}{\xi} \nabla r_n' + \nabla r_n \frac{\nabla \xi'}{\xi} + \nabla^2 r_n + \nabla r_n \nabla r_n'\right) Q_n'\right]\right\}$$

and

$$\mathrm{IV}_n(h) = E_\xi^n\left\{\mathrm{tr}\left[Vh(Z_n)Q_n\left(\frac{\nabla^2\xi}{\xi}\right)Q_n' - (\Phi^p Vh)Q_\theta\left(\frac{\nabla^2\xi}{\xi}\right)Q_\theta'\right]\right\}$$

and the dependence of $\nabla\xi$ and ∇r_n on θ has been suppressed in the notation. The terms involving I_n–IV_n are shown to be negligible compared to $1/n$ below, and it follows that:

$$E_\xi\{h(Z_n)\} = \Phi^p h + \frac{1}{\sqrt{n}}(\Phi^p Uh)' E_\xi\left\{Q_\theta\left(\frac{\nabla\xi}{\xi}\right)\right\} + \frac{1}{n}E_\xi\left\{\mathrm{tr}\left[(\Phi^p Vh)Q_\theta\left(\frac{\nabla^2\xi}{\xi}\right)Q_\theta'\right]\right\} + \mathrm{o}\left(\frac{1}{n}\right). \quad (14)$$

In fact, (14) holds for all $h \in \mathscr{H}_2$ and uniformly with respect to $h \in \mathscr{H}_2^o$.

Write $Q_\theta = [q_{ij}(\theta) : i = 1, \ldots, p, j = 1, \ldots, p]$ for $\theta \in \Re^p$, and let $Q_\theta^\# = [q_{ij}^\#(\theta) : i, j = 1, \ldots, p]$ and $M_\theta = [m_{ij}(\theta) : i, j = 1, \ldots, p]$, where

$$q_{ij}^\#(\theta) = \frac{\partial q_{ij}(\theta)}{\partial\theta_j} \quad (15)$$

and

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

Integrating by parts,

$$E_\xi\left\{Q_\theta\left(\frac{\nabla\xi}{\xi}\right)\right\} = -\int (Q_\theta^\#\mathbf{1})\xi(\theta)\,\mathrm{d}\theta$$

and

$$E_\xi\left\{\mathrm{tr}\left[(\Phi^p Vh)Q_\theta\left(\frac{\nabla^2\xi}{\xi}\right)Q_\theta'\right]\right\} = \int \mathrm{tr}[(\Phi^p Vh)M_\theta]\xi(\theta)\,\mathrm{d}\theta,$$

where $\mathbf{1} = [1, \ldots, 1]'$. So, (14) becomes

$$E_\xi\{h(Z_n)\} = \int\left\{\Phi^p h - \frac{1}{\sqrt{n}}(\Phi^p Uh)' Q_\theta^\#\mathbf{1} + \frac{1}{n}\mathrm{tr}[(\Phi^p Vh)M_\theta]\right\}\xi(\theta)\,\mathrm{d}\theta + \mathrm{o}\left(\frac{1}{n}\right)$$

or

$$E_\theta\{h(Z_n)\} = \Phi^p h - \frac{1}{\sqrt{n}}(\Phi^p Uh)' Q_\theta^\#\mathbf{1} + \frac{1}{n}\mathrm{tr}[(\Phi^p Vh)M_\theta] + \mathrm{o}\left(\frac{1}{n}\right) \quad (16)$$

very weakly. The forms of asymptotic expansions in (14) and (16) agree with those in Woodroofe and Coad (1997), but the definitions of $\hat{\theta}_n$, B_n, and Z_n are different.

If $h(z) = z$, then $\Phi^p h = 0$ and $Uh(z) = I_p = \Phi^p Uh$. Applying (16) to this h suggests

$$E_\theta(Z_n) \approx -\frac{1}{\sqrt{n}}\, Q_\theta^\#\mathbf{1} = \mu_n(\theta), \quad (17)$$

Let

$$\hat{\mu}_{ni} = \begin{cases} -\sum_{j=1}^{p} q_{ij}^{\#}(\hat{\theta}_n)/\sqrt{n} & \text{if } |\sum_{j=1}^{p} q_{ij}^{\#}(\hat{\theta}_n)| \leqslant \sqrt{n} \\ -\mathrm{sgn}[\sum_{j=1}^{p} q_{ij}^{\#}(\hat{\theta}_n)] & \text{otherwise} \end{cases} \tag{18}$$

for $i = 1, \ldots, p$, $\hat{\mu}_n = (\hat{\mu}_{n1}, \ldots, \hat{\mu}_{np})'$, and consider $(Z_n - \hat{\mu}_n)$. Approximations like those described above lead to

$$E_\theta\{(Z_n - \hat{\mu}_n)(Z_n - \hat{\mu}_n)'\} = I_p + \frac{\Delta(\theta)}{n} + \mathrm{o}\left(\frac{1}{n}\right)$$

very weakly, where $\Delta(\theta) = [\delta_{ij}(\theta) : i, j = 1, \ldots, p]$ and

$$\delta_{ij}(\theta) = \sum_{k=1}^{p} \sum_{l=1}^{p} \left(\frac{\partial q_{ik}}{\partial \theta_l}\right)\left(\frac{\partial q_{jl}}{\partial \theta_k}\right). \tag{19}$$

Next, let $\hat{\delta}_{n,ij} = \delta_{ij}(\hat{\theta}_n)$ if $|\delta_{ij}(\hat{\theta}_n)| \leqslant n$, $\hat{\delta}_{ij} = 0$ otherwise, and $\hat{\Delta}_n = [\hat{\delta}_{n,ij} : i, j = 1, \ldots, p]$; and let $\hat{\Gamma}_n$ be any (measurable) $p \times p$ matrices for which

$$\lim_{n\to\infty} nE_\xi \left\| \left(\frac{\hat{\Gamma}_n + \hat{\Gamma}_n'}{2}\right) - \left[I_p + \frac{\hat{\Delta}_n}{2n}\right] \right\| = 0 \tag{20}$$

for any ξ (under consideration). The choice $\hat{\Gamma}_n = (I_p + \hat{\Delta}_n/2n)$ always satisfies (20), but other choices may be convenient in applications. The main result asserts that (6) and (7) hold with these choices of $\hat{\mu}_n$ and $\hat{\Gamma}_n$. This will be proved in Section 6.

4. An example

In this section we compare the theoretical results to simulation experiments. Consider an AR(2) process,

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

where e_t are independent standard normal random variables, and $\theta = (\theta_1, \theta_2)' \in \Omega$ and $\sigma > 0$ are unknown parameters. For $\{y_t\}$ to be causal, the parameter space Ω is determined by the inequalities: $\theta_1 + \theta_2 < 1$, $\theta_1 - \theta_2 > -1$, and $\theta_2 > -1$. See Brockwell and Davis (1991, Chapter 3). These inequalities imply $|\theta_2| < 1$. When $\sigma = 1$, the covariance matrix G_θ has inverse

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

The procedure for setting confidence intervals is illustrated for θ_2; the treatment of θ_1 is similar. If B_n is a lower triangular matrix in (2) and $\hat{\Gamma}_n$ is an upper triangular matrix for which $\hat{\Gamma}_n \hat{\Gamma}_n' = I_p + \hat{\Delta}_n / n$, then (20) holds, and

$$\theta_2 - \hat{\theta}_{n,2} = \frac{\tilde{\sigma}_n}{b_n} \left\{ \hat{\mu}_{n,2} + \sqrt{\left(1 + \frac{\hat{\delta}_{n,22}}{n}\right)} \times T^*_{n,2} \right\},$$

where b_n is the lower right-hand entry in B_n and $T^*_{n,2}$ is the second component of T^*_n. With this choice of B_n,

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

So, by (15), (17) and (19),

$$\mu_{n,2}(\theta) = \frac{1+2\theta_2}{\sqrt{n(1-\theta_2^2)}} \tag{21}$$

and

$$\delta_{22} = \frac{(1+2\theta_2)^2}{1-\theta_2^2}.$$

Since T^*_{n2} is asymptotically t_n to order $o(1/n)$, an asymptotic level γ confidence interval for θ_2 is $\{|T^*_{n,2}| \leqslant c_n\}$, where c_n is the $100(1+\gamma)/2$ quantile of the standard univariate t-distribution with n degrees of freedom, i.e.

$$\hat{\theta}_{n,2} + \frac{\tilde{\sigma}_n}{b_n} \hat{\mu}_{n,2} \pm \frac{\tilde{\sigma}_n}{b_n} \sqrt{\left(1 + \frac{\hat{\delta}_{n,22}}{n}\right)} \times c_n.$$

Table 1 reports the simulated values of $P_\theta(T_{n2} \geqslant 2.228)$, $P_\theta(T_{n2} \leqslant -2.228)$, $P_\theta(|T_{n2}| \leqslant 2.228)$, $E_\theta(T_{n2})$, and $E_\theta(T_{n2}^2)$, for $\sigma = 1$ and $n = 10$; and similarly for T^*_{n2}. Here -2.228 is the 2.5th percentile of the standard univariate t-distribution with 10 degrees of freedom. The notation $\pm$ in the last row indicates 1.96 standard deviations; for example, ± 0.022 is obtained by $E(t_{10}) \pm 1.96 \times [\mathrm{Var}(t_{10})/10{,}000]^{1/2}$, 1.25 ± 0.042 is by $E(t_{10}^2) \pm 1.96 \times [\mathrm{Var}(t_{10}^2)/10{,}000]^{1/2}$, etc. Results for $n = 20$ and 50 are given in Tables 2 and 3, respectively. For T_{n2}, the simulated values of $P_\theta(T_{n2} \geqslant c_n)$ and $P_\theta(T_{n2} \leqslant -c_n)$ are not sensitive to θ_1, but quite sensitive to θ_2. For $\theta_2 = 0.0, 0.5$, these values are significantly different from the nominal value 0.025 at significance level 0.05 even for $n = 50$; for $\theta_2 = -0.5$ and $n = 50$, they agree well with the nominal value 0.025. For all choices of n, the values of $E_\theta(T_{n2})$ are similarly not sensitive to θ_1, but sensitive to θ_2. They are significantly different from zero for $\theta_2 \neq -0.5$ at significance level 0.05. All these features can be explained from (21), which says that the theoretical mean of T_{n2} depends only on θ_2 and it vanishes when $\theta_2 = -0.5$. For the refined pivot T^*_{n2}, the simulated values of $P_\theta(T^*_{n2} \geqslant c_n)$ and $P_\theta(T^*_{n2} \leqslant -c_n)$ show that T^*_{n2} is not symmetric in the tails. Especially, for $n = 10$ and $\theta_2 = 0.5$, these coverage

Table 1
$n = 10$, replicates $= 10,000$; $c_n = 2.228$; $\pm$ is the range within 1.96 standard deviations

(θ_1, θ_2)	$E_\theta(T_{n2})$ $E_\theta(T^*_{n2})$	$E_\theta(T^2_{n2})$ $E_\theta(T^{*2}_{n2})$	$P_\theta(T_{n2} \geqslant c_n)$ $P_\theta(T^*_{n2} \geqslant c_n)$	$P_\theta(T_{n2} \leqslant -c_n)$ $P_\theta(T^*_{n2} \leqslant -c_n)$	$P_\theta(\lvert T_{n2}\rvert \leqslant c_n)$ $P_\theta(\lvert T^*_{n2}\rvert \leqslant c_n)$
(0.0 −0.5)	−0.012	1.005	0.016	0.016	0.967
	−0.014	1.130	0.024	0.023	0.953
(0.0 0.0)	0.281	1.075	0.034	0.011	0.955
	0.095	1.494	0.037	0.005	0.958
(0.0 0.5)	0.552	1.196	0.053	0.008	0.939
	0.461	3.568	0.048	0.000	0.952
(0.5 −0.5)	−0.020	1.008	0.016	0.016	0.968
	−0.029	1.132	0.022	0.023	0.955
(0.5 −0.2)	0.148	1.050	0.025	0.013	0.962
	0.026	1.215	0.030	0.016	0.954
(0.5 0.0)	0.250	1.100	0.032	0.014	0.954
	0.077	1.409	0.035	0.005	0.960
$\pm$	$\pm$0.022	1.25 $\pm$ 0.042	0.025 $\pm$ 0.003	0.025 $\pm$ 0.003	0.95 $\pm$ 0.004

Table 2
$n = 20$, replicates $= 10,000$; $c_n = 2.086$; $\pm$ is the range within 1.96 standard deviations

(θ_1, θ_2)	$E_\theta(T_{n2})$ $E_\theta(T^*_{n2})$	$E_\theta(T^2_{n2})$ $E_\theta(T^{*2}_{n2})$	$P_\theta(T_{n2} \geqslant c_n)$ $P_\theta(T^*_{n2} \geqslant c_n)$	$P_\theta(T_{n2} \leqslant -c_n)$ $P_\theta(T^*_{n2} \leqslant -c_n)$	$P_\theta(\lvert T_{n2}\rvert \leqslant c_n)$ $P_\theta(\lvert T^*_{n2}\rvert \leqslant c_n)$
(0.0 −0.5)	−0.005	0.994	0.019	0.020	0.961
	−0.006	1.075	0.026	0.027	0.947
(0.0 0.0)	0.215	1.049	0.037	0.012	0.951
	0.030	1.084	0.034	0.020	0.946
(0.0 0.5)	0.449	1.105	0.054	0.006	0.939
	0.081	1.011	0.036	0.000	0.964
(0.5 −0.5)	−0.003	0.987	0.018	0.018	0.964
	−0.003	1.068	0.025	0.026	0.949
(0.5 −0.2)	0.127	1.022	0.028	0.015	0.957
	0.012	1.082	0.030	0.024	0.946
(0.5 0.0)	0.207	1.044	0.035	0.013	0.953
	0.022	1.078	0.032	0.020	0.949
$\pm$	$\pm$0.021	1.11 $\pm$ 0.033	0.025 $\pm$ 0.003	0.025 $\pm$ 0.003	0.95 $\pm$ 0.004

probabilities indicate that the distribution of T^*_{n2} is skewed to the right. Correspondingly, the estimated values of $E_\theta(T^*_{n2})$ and $E_\theta(T^{*2}_{n2})$ are far above their nominal values. This skewness may be due to the facts that $\hat{\theta}_{n,2}$ tends to under-estimate θ_2, and that from (21) the downward bias of $\hat{\mu}_{n,2}$ is more severe for larger θ_2. The tables also show that the estimated means of T^*_{n2} are much closer to 0 than those of T_{n2}, and for $n=50$ the values of $E_\theta(T^*_{n2})$ are all within 1.96 standard deviations. Although the difference between the simulated values of $E_\theta(T^{*2}_{n2})$ and $E_\theta(T^2_{n2})$ are less obvious, in general the former is larger and closer to

Table 3
$n = 50$, replicates = 10, 000; $c_n = 2.008$; ± is the range within 1.96 standard deviations

(θ_1, θ_2)	$E_\theta(T_{n2})$ $E_\theta(T^*_{n2})$	$E_\theta(T^2_{n2})$ $E_\theta(T^{*2}_{n2})$	$P_\theta(T_{n2} \geqslant c_n)$ $P_\theta(T^*_{n2} \geqslant c_n)$	$P_\theta(T_{n2} \leqslant -c_n)$ $P_\theta(T^*_{n2} \leqslant -c_n)$	$P_\theta(\|T_{n2}\| \leqslant c_n)$ $P_\theta(\|T^*_{n2}\| \leqslant c_n)$
(0.0 −0.5)	−0.013	0.999	0.023	0.023	0.954
	−0.014	1.036	0.027	0.027	0.946
(0.0 0.0)	0.127	1.026	0.033	0.019	0.948
	−0.005	1.045	0.029	0.028	0.943
(0.0 0.5)	0.303	1.046	0.045	0.011	0.943
	0.015	1.003	0.026	0.021	0.953
(0.5 −0.5)	−0.003	0.989	0.023	0.021	0.957
	−0.003	1.026	0.026	0.024	0.950
(0.5 −0.2)	0.083	1.018	0.029	0.020	0.951
	0.002	1.046	0.028	0.026	0.946
(0.5 0.0)	0.138	1.031	0.033	0.018	0.949
	0.006	1.049	0.029	0.025	0.945
±	±0.020	1.04 ± 0.030	0.025 ± 0.003	0.025 ± 0.003	0.95 ± 0.004

the nominal values. The implementation of simulations is written in C; it is available at http://www3.nccu.edu.tw/~chweng/AR2.c.

5. Some bounds

Several bounds are needed for the proofs of the main results. Throughout this section, ξ denotes a twice continuously differentiable density with compact, convex support $K \subseteq \Omega$, and $\|A\|$ denotes the spectral norm of a matrix A, i.e. $\|A\|^2 = \lambda_{\max}(A'A)$.

Lemma 5.2. $\inf_{\theta\in\Omega} \lambda_{\min} G_\theta > 0$ *and* $\lim_{\theta\to\partial\Omega} \det G_\theta = \infty$.

Proof. By Brockwell and Davis (1991, 137pp), $\lambda_{\min}(G_\theta) \geqslant 2\pi \inf_\lambda f_\theta(\lambda)$, where $f_\theta(\lambda)$ is the spectral density of $\{y_t\}$. For an AR(p) process, $f_\theta(\lambda) = 1/[2\pi|1 - \sum_{j=1}^p \theta_j e^{-ij\lambda}|^2]$, which is bounded below for all $\theta \in \Omega$. Next, let $v = (1, 0, \ldots, 0)'$. Then

$$\lambda_{\max} G_\theta \geqslant v' G_\theta v = \int_{-\pi}^{\pi} f_\theta(\lambda)\, d\lambda \to \infty$$

as θ approaches any boundary point of Ω. □

Lemma 5.3. *For every* $s \geqslant 1$, *there is an integer* n_s *and a continuous function* $C_s(\theta)$ *for which*

$$E_\theta\left\{\frac{n}{\lambda_{\min}(X_n' X_n)^s}\right\} \leqslant C_s(\theta)$$

for all $n \geqslant n_s$ *and all* $\theta \in \Omega$.

Proof. Let $x_i = (y_{i+p-1}, \ldots, y_i)'$, $i = 1, \ldots, n$. Further, let $q = p + 2s$ and let $r = r_n$ be positive integers for which $pqr \leqslant n < pq(r+1)$. Then

$$X_n' X_n = \sum_{i=1}^{n} x_i x_i' \geqslant \sum_{k=1}^{r} \sum_{j=1}^{q} x_{pq(k-1)+p(j-1)+1} x'_{pq(k-1)+p(j-1)+1} = \sum_{k=1}^{r} N_k,$$

in the sense of positive definite matrices. So, $\lambda_{\min}(X_n' X_n) \geqslant \sum_{k=1}^{r} \lambda_{\min}(N_k)$; and, since $f(x) = 1/x^s$ is convex in $0 < x < \infty$,

$$\frac{n^s}{(\lambda_{\min}(X_n' X_n))^s} \leqslant \left(\frac{n}{r}\right)^s \frac{1}{\left[\frac{1}{r}\sum_{k=1}^{r} \lambda_{\min}(N_k)\right]^s} \leqslant (2pq)^s \frac{1}{r} \sum_{k=1}^{r} \frac{1}{\lambda_{\min}(N_k)^s}$$

and

$$E_\theta \left\{ \frac{n^s}{\lambda_{\min}(X_n' X_n)^s} \right\} \leqslant (2pq)^s E_\theta \left\{ \frac{1}{\lambda_{\min}(N_1)^s} \right\}.$$

Observe that $N_1 = x_1 x_1' + x_{p+1} x'_{p+1} + \cdots + x_{p(q-1)+1} x'_{p(q-1)+1}$ is a function of $U = (y_1, \ldots, y_{pq})'$, say $N_1 = g(U)$. Here U is normally distributed with mean zero and covariance matrix Γ. Let γ denote the maximum eigenvalue of Γ. Let $W = (w_1, \ldots, w_{pq})'$ be a pq-variate normally distributed random vector with mean zero and covariance matrix γI_{pq}. Then $f_U \leqslant c_\theta f_W$, where f_U and f_W are density functions of U and W, and $c_\theta = \{\det(\Gamma)\}^{-1/2} \gamma^{pq/2}$ depends continuously on θ. Now $g(W)$ follows the Wishart distribution with q degrees of freedom, dimension p and covariance matrix γI_p. For $q \geqslant p$, the joint density of the eigenvalues of $g(W)$, $l_1 > \cdots > l_p > 0$, is

$$f(l_1, \ldots, l_p) = C e^{-1/2\gamma \sum_{i=1}^{p} l_i} \prod_{i=1}^{p} l_i^{(q-p-1)/2+p-i} \times \prod_{i<j} (l_i - l_j).$$

With $q = p + 2s$, $(q - p - 1)/2 - s = -\frac{1}{2}$ and

$$E_\theta \left\{ \left(\frac{1}{\lambda_{\min}(N_1)} \right)^s \right\} = \int \left[\frac{1}{\lambda_{\min}(g(u))} \right]^s F_U(\mathrm{d}u)$$

$$\leqslant c_\theta \int \cdots \int \frac{1}{l_p^s} f(l_1, \ldots, l_p) \, \mathrm{d}l_p \cdots \mathrm{d}l_1,$$

which is finite, and the lemma follows with $n_s = p(p + 2s)$. □

Recall that $\tilde{\theta}_n$ and $\tilde{\sigma}_n^2$ denote the least squares estimators, $\tilde{\theta}_n = (X_n' X_n)^{-1} X_n' y_{p,n}$ and $\tilde{\sigma}_n^2 = \|y_{p,n} - X_n \tilde{\theta}_n\|^2/(n-p)$; let $\tilde{B}_n \tilde{B}_n' = X_n' X_n$ be the Cholesky decomposition of $X_n' X_n$, and let $\tilde{Z}_n = \tilde{B}_n'(\theta - \tilde{\theta}_n)$. Then $\|\tilde{Z}_n\|^2 = (\theta - \tilde{\theta}_n)' X_n' X_n (\theta - \tilde{\theta}_n)$. It is known and clear that $\left[\sqrt{n}(\tilde{\sigma}_n^2 - 1), \tilde{Z}_n\right] \Rightarrow [U, Z]$ for fixed θ, where $U \sim N[0, 2]$ and $Z \sim N_p[0, I_p]$ are independent and it follows easily that $\left[\theta, \sqrt{n}(\tilde{\sigma}_n^2 - 1), \tilde{Z}_n\right] \Rightarrow [\theta, U, Z]$, under P_ξ.

Lemma 5.4. *For any* $k \geqslant 1$, $\|\tilde{Z}_n\|^{2k}$ *and* $n^k(\tilde{\sigma}_n^2 - 1)^{2k}$, $n > p$, *are uniformly integrable w.r.t.* P_ξ. *Moreover*, $E_\xi(\tilde{\sigma}_n^2 - 1) = \mathrm{o}(1/n)$.

Proof. For the uniform integrability of $\|\tilde{Z}_n\|^{2k}$, $n > p$, first observe that $\|\tilde{Z}_n\|^2 = e'_{p,n}X_n(X'_n X_n)^{-1}X'_n e_{p,n} \leqslant p\|e_{p,n}\|^2$, so that $E_\xi\|\tilde{Z}_n\|^{2k} = \mathrm{O}(n^k)$ for every k. Next, let $A_n = \{\lambda_{\min}(X'_n X_n) \geqslant 1\}$ and $g_k(z) = 1 + \|z\|^{2k}$. Then

$$E_\xi\|\tilde{Z}_n\|^{2k}1_{A_n^c} \leqslant \sqrt{E_\xi\|\tilde{Z}_n\|^{4k}}\sqrt{P(A_n^c)} \to 0 \quad \text{as } n \to \infty$$

for every k, by Lemma 5.3. So, it suffices to show that $\sup_{n>p} E_\xi[g_k(Z_n)1_{A_n}] < \infty$ for every $k \geqslant 1$. Now $\Phi^p U g_k = 0$, and there are constants C_k for which $\|Vg_k(z)\| \leqslant C_k g_{k-1}$. So,

$$E_\xi^n[g_k(\tilde{Z}_n)\mathbf{1}_{A_n}] = \Phi g_k \mathbf{1}_{A_n} + E_\xi^n\left\{\mathrm{tr}\left[Vg_k(\tilde{Z}_n)\tilde{B}_n^{-1}\frac{\nabla^2\xi}{\xi}\tilde{B}_n'^{-1}1_{A_n}\right]\right\}$$

$$\leqslant \Phi g_k + C_k E_\xi^n\left[g_{k-1}(\tilde{Z}_n)\|\tilde{B}_n^{-1}\|^2\left\|\frac{\nabla^2\xi}{\xi}\right\|1_{A_n}\right]$$

and therefore,

$$E_\xi[g_k(\tilde{Z}_n)1_{A_n}] \leqslant \Phi g_k + C_k\int_\Omega E_\theta[g_{k-1}(\tilde{Z}_n)1_{A_n}]\|\nabla^2\xi(\theta)\|\,\mathrm{d}\theta. \tag{22}$$

There is a second twice continuously differentiable density $\tilde{\xi}$ with compact convex support and a constant C for which $\|\nabla^2\xi(\theta)\| \leqslant C\tilde{\xi}(\theta)$ for all θ. Then the right side of (22) is at most $\Phi g_k + CC_k E_{\tilde{\xi}}[g_{k-1}(\tilde{Z}_n)\mathbf{1}_{A_n}]$. That $\sup_{n>p} E_\xi[g_k(\tilde{Z}_n)1_{A_n}] < \infty$ then follows by induction over k. The assertions concerning $\tilde{\sigma}_n^2$ then follow from

$$\tilde{\sigma}_n^2 - 1 = \frac{\|y_{p,n} - X_n\theta\|^2 - n - (\|\tilde{Z}_n\|^2 - p)}{n - p},$$

since $\|y_{p,n} - X_n\theta\|^2 \sim \chi_n^2$. □

Lemma 5.5. *For any* $k \geqslant 1$, $n^k(\hat{\sigma}_n^2 - 1)^{2k}$, $n > p$, *are uniformly integrable with respect to* P_ξ. *Further*, $\left[\theta, y_{0,p}, \sqrt{n}(\hat{\sigma}_n^2 - 1)\right] \Rightarrow [\theta, y_{0,p}, U]$, *where* $U \sim \mathrm{Normal}[0, 2]$ *is independent of* $[\theta, y_{0,p}]$.

Proof. From (1) and the non-negativity of $y'_{0,p}G_{\hat{\theta}_n}^{-1}y_{0,p}$,

$$\hat{\sigma}_n^2 \geqslant \frac{\|y_{p,n} - X_n\tilde{\theta}_n\|^2}{n+p} = \frac{\|y_{p,n} - X_n\theta\|^2 - \|\tilde{Z}_n\|^2}{n+p}.$$

Next, from (1) and the likelihood function,

$$\begin{aligned}(n+p)\hat{\sigma}_n^2 &= -2\hat{\sigma}_n^2\ell_n(\hat{\sigma}_n^2,\hat{\theta}_n) - \hat{\sigma}_n^2\log|G_{\hat{\theta}_n}| - (n+p)\hat{\sigma}_n^2\log(\hat{\sigma}_n^2)\\ &\leqslant -2\hat{\sigma}_n^2\ell_n(\hat{\sigma}_n^2,\theta) - \hat{\sigma}_n^2\log|G_{\hat{\theta}_n}| - (n+p)\hat{\sigma}_n^2\log(\hat{\sigma}_n^2)\\ &= \|y_{p,n} - X_n\theta\|^2 + y_{0,p}'G_\theta^{-1}y_{0,p} + \hat{\sigma}_n^2\log|G_\theta| - \hat{\sigma}_n^2\log|G_{\hat{\theta}_n}|\\ &\leqslant \|y_{p,n} - X_n\theta\|^2 + y_{0,p}'G_\theta^{-1}y_{0,p} + C\hat{\sigma}_n^2\end{aligned}$$

for $\theta\in K$ and some constant C depending on ξ. Let $U_n = \|y_{p,n} - X_n\theta\|^2$, so that $U_n\sim\chi_n^2$ is independent of $y_1,\ldots,y_p$. Then

$$\frac{U_n - n - \|\tilde{Z}_n\|^2 - p}{n+p}\leqslant\hat{\sigma}_n^2 - 1\leqslant\frac{U_n - n + y_{0,p}'G_\theta^{-1}y_{0,p} - p + C}{n+p-C}$$

for $n > C - p$. The lemma follows from these inequalities and Lemma 5.4. □

Recall that ξ has a compact convex support K and that $\lim_{\theta\to\partial\Omega}\det G_\theta = \infty$, from Lemma 5.2. So, there exist two other compact convex sets K_1 and K_2 for which $K\subset K_1^0\subset K_1\subset K_2^0\subset K_2\subset\Omega$, where K_i^0 denotes the interior of K_i, and $\det[G_\theta]\geqslant 1$ for all $\theta\notin K_1$.

Lemma 5.6. *There are events D_n for which*

$$D_n\supseteq\{\tfrac{1}{2}\leqslant\hat{\sigma}_n^2\leqslant\tfrac{3}{2},\ \tilde{\theta}\in K_1,\ \hat{\theta}\in K_2,\ \|y_{0,p}\|\leqslant n^{1/4},\ \lambda_{\min}(X_n'X_n)\geqslant n^{3/4}\}\tag{23}$$

$$P_\xi(D_n^c) = \mathrm{o}\left(\frac{1}{n^k}\right)\tag{24}$$

for every $k\geqslant 1$.

Proof. To show existence, let D_n be the right side of (23). Then it suffices to show (24). For this choice of D_n,

$$\begin{aligned}D_n^c &\subseteq\{|\hat{\sigma}_n^2 - 1|\geqslant\tfrac{1}{2}\}\cup\{\|y_{0,p}\| > n^{1/4}\}\cup\{\lambda_{\min}(X_n'X_n) < n^{3/4}\}\\ &\quad\cup\{\|y_{0,p}\|\leqslant n^{1/4},\ \lambda_{\min}(X_n'X_n)\geqslant n^{3/4},\ \tilde{\theta}_n\notin K_1\}\\ &\quad\cup\{\|y_{0,p}\|\leqslant n^{1/4},\ \lambda_{\min}(X_n'X_n)\geqslant n^{3/4},\ \tilde{\theta}_n\in K_1,\ \hat{\theta}_n\notin K_2\}\\ &= A_1\cup A_2\cup A_3\cup A_4\cup A_5.\end{aligned}\tag{25}$$

It is clear from Lemmas 5.3 and 5.5 that $P_\xi(A_1) + P_\xi(A_2) + P_\xi(A_3) = \mathrm{o}(1/n^k)$ for every $k\geqslant 1$. Let $\delta_1 = \mathrm{dist}(K, K_1^c) > 0$. If A_4 occurs, then $\|\tilde{\theta} - \theta\|\geqslant\delta_1$ and therefore, $\|\tilde{Z}_n\|^2 = (\tilde{\theta} - \theta)'X_n'X_n(\tilde{\theta} - \theta)\geqslant\delta_1^2 n^{3/4}$. So,

$$P_\xi(A_4)\leqslant\frac{1}{\delta_1^{2k}n^{3/2k}}E_\xi\|\tilde{Z}_n\|^{4k} = \mathrm{o}\left(\frac{1}{n^k}\right)$$

for all k. For A_5, let $\delta_2 = \mathrm{dist}(K_1, K_2^c)$ and observe that if A_5 occurs, then $\|X_n(\hat{\theta}_n - \tilde{\theta}_n)\|^2 \geqslant \delta_2^2 n^{3/4}$. If $\tilde{\theta}_n \in K_1$, then using the form of the likelihood function and Lemma 5.2,

$$\begin{aligned}\|X_n(\hat{\theta}_n - \tilde{\theta}_n)\|^2 &= \|y_{p,n} - X_n\hat{\theta}_n\|^2 - \|y_{p,n} - X_n\tilde{\theta}_n\|^2 \\ &= -2\hat{\sigma}_n^2 \ell(\hat{\sigma}_n^2, \hat{\theta}_n) - y_{0,p}' G_{\hat{\theta}_n}^{-1} y_{0,p} - \hat{\sigma}_n^2 \log[\det(G_{\hat{\theta}_n})] \\ &\quad + 2\hat{\sigma}_n^2 \ell(\hat{\sigma}_n^2, \tilde{\theta}_n) + y_{0,p}' G_{\tilde{\theta}_n}^{-1} y_{0,p} + \hat{\sigma}_n^2 \log[\det(G_{\tilde{\theta}_n})] \\ &\leqslant 2C_0 + 2C_1 + C_2\|y_{0,p}\|^2, \qquad (26)\end{aligned}$$

where $-C_0$ denotes a lower bound for $\log[\det(G_\omega)]$, $\omega \in \Omega$, C_1 denotes an upper bound for $\log[\det(G_\omega)]$, $\omega \in K_1$, and $1/C_2$ denotes a lower bound for $\lambda_{\min}[G_\omega]$, $\omega \in K_1$. So, A_5 is empty for sufficiently large n. □

Lemma 5.7. *For any $k \geqslant 1$, $\|X_n(\hat{\theta}_n - \theta)\|^{2k}$ and $n^k\|\hat{\theta}_n - \theta\|^{2k}$ are uniformly integrable.*

Proof. Since $\hat{\theta}_n$ and θ are bounded and $E_\xi\|X_n'X_n\|^{2k} = \mathrm{O}(n^{2k})$ for any k, it suffices to show that $\|Z_n\|^{2k}1_{D_n}$ and $n^k\|\theta - \hat{\theta}_n\|^{2k}1_{D_n}$ are uniformly integrable for events D_n that satisfy (23) and (24). Uniform integrability of $\|Z_n\|^{2k}1_{D_n}$ follows directly from Lemma 5.4 and (26); and that of $n\|\hat{\theta}_n - \theta\|^2\mathbf{1}_{D_n}$ then follows from

$$n\|\theta - \hat{\theta}\|^2 1_{D_n} \leqslant \frac{n}{\lambda_{\min}(X_n'X_n)}(\theta - \hat{\theta})'X_n'X_n(\theta - \hat{\theta})1_{D_n},$$

Lemma 5.3, Schwarz's Inequality, and the first assertion. □

Recall the definition of r_n from (8).

Proposition 5.8. *Let D_n satisfy* (23) *and let* $2\kappa_\theta = \log[\det(G_\theta)]$. *Then*

(a) *For any $q \geqslant 1$, $(\|\nabla r_n\|^2 + \|\nabla^2 r_n\|)^q 1_{D_n}$ are uniformly integrable.*
(b) *$n\|\nabla r_n\|^2 1_{D_n}$ are uniformly integrable.*
(c) $\left[\theta, \sqrt{n}\nabla r_n(\theta)\right] \Rightarrow [\theta, \nabla\kappa_\theta U]$, *where* $U \sim \mathrm{Normal}[0, 2]$ *is independent of* $[\theta, y_{0,p}]$.

Proof. Again using the form of the likelihood function, $\ell_n(\theta) - \sigma^2\ell_n(\sigma^2, \theta) = \ell_0(\theta) - \sigma^2\ell_0(\sigma^2, \theta) + n\sigma^2\log\sigma$, so that

$$\nabla r_n(\hat{\theta}_n) = \nabla\ell_n(\hat{\theta}_n) = \nabla\ell_0(\hat{\theta}_n) - \hat{\sigma}_n^2\nabla\ell_0(\hat{\sigma}_n^2, \hat{\theta}_n) = (\hat{\sigma}_n^2 - 1)(\nabla\kappa_{\hat{\theta}_n}),$$

$$\nabla^2 r_n(\theta) = \nabla^2\ell_0(\theta) - \hat{\sigma}_n^2\nabla^2\ell_0(\hat{\sigma}_n^2, \hat{\theta}_n) = \mathrm{o}_p(1)$$

and

$$\nabla r_n(\theta) = \nabla r_n(\hat{\theta}_n) + \int_0^1 \nabla^2 r_n[t\theta + (1-t)\hat{\theta}_n](\theta - \hat{\theta}_n)\,\mathrm{d}t.$$

The assertion (c) follows immediately. Next, since $\hat{\sigma}_n$ and $\hat{\theta}_n$ are bounded on D_n, there is a constant C for which

$$\|\nabla r_n(\theta)\|^2 \leqslant C[(\hat{\sigma}_n^2 - 1)^2 + \|\hat{\theta}_n - \theta\|^2]$$

on D_n. Assertions (a) and (b) follow directly from Lemmas 5.5–5.7. □

Proposition 5.9. *Let D_n be as in* (23) *and* (24). *Then*,

(a) *for any $q \geqslant 1$, $\|Q_n\|^{2q} 1_{D_n}$, $n \geqslant n_q$, are uniformly integrable w.r.t. P_ξ*,
(b) $\lim_{n\to\infty} \sqrt{n} \int_\Omega \|E_\theta[(Q_n - Q_\theta)1_{D_n}]\| \xi(\theta)\, \mathrm{d}\theta = 0$.

Proof. For (a), note that on D_n we have $\hat{\theta} \in K_2$. So, there exist constants C_0 and C_1 such that $\|\nabla^2 \ell_0(\hat{\sigma}_n^2, \hat{\theta}_n)\| \leqslant C_0 + C_1\|y_{0,p}\|^2 \leqslant C_0 + C_1 n^{1/2}$ on D_n. It follows that

$$\begin{aligned}\|Q_n\|^2 1_{D_n} &= \lambda_{\max}[n(X_n' X_n - \hat{\sigma}_n^2 \nabla^2 \ell_0(\hat{\sigma}_n^2, \hat{\theta}_n))^{-1}] 1_{D_n} \\ &\leqslant \frac{n}{\lambda_{\min}(X_n' X_n - 2\nabla^2 \ell_0(\hat{\sigma}_n^2, \hat{\theta}_n))} 1_{D_n} \\ &\leqslant \frac{2n}{\lambda_{\min}(X_n' X_n)} 1_{D_n}\end{aligned}$$

for sufficiently large n. Assertion (a) follows from Lemma 5.3 and the compactness of K.

For (b) if M is a $p \times p$ matrix, denote vec(M) as the $p^2 \times 1$ vector composed of the p row vectors. Write the ij component of Q_n as $Q_n^{ij} = g(w_n + s_n)$, where g is a smooth function and

$$w_n = \mathrm{vec}\left(\frac{1}{n} X_n' X_n\right) \quad \text{and} \quad s_n = \mathrm{vec}\left(-\frac{1}{n}\hat{\sigma}_n^2 \nabla^2 \ell_0(\hat{\sigma}_n^2, \hat{\theta}_n)\right)$$

are p^2-variate random vectors. Since $s_n \to 0$ $w.p.1$ and $w_n \to \mu = \mathrm{vec}(G_\theta) w.p.1$, we have $g(\mu) = Q_\theta^{ij}$. Then an application of Taylor's expansion leads to

$$Q_n^{ij} = Q_\theta^{ij} + \nabla g(\mu)'(w_n + s_n - \mu) + \tfrac{1}{2}(w_n + s_n - \mu)' \nabla^2 g(\delta)(w_n + s_n - \mu),$$

where δ lies between μ and $w_n + s_n$. Let $B_{\mu;\eta}$ denote the ball centered at μ with radius $\eta > 0$ and define $A_n = \{w_n \in B_{\mu;\eta/2}$ and $\|s_n\| \leqslant \eta/2\}$. Write

$$\begin{aligned}E_\xi(Q_n^{ij} 1_{D_n}) - E_\xi(Q_\theta^{ij} 1_{D_n}) &= E_\xi((Q_n^{ij} - Q_\theta^{ij}) 1_{D_n \cap A_n^c}) \\ &\quad + E_\xi\{\nabla g(\mu)'(w_n + s_n - \mu) 1_{D_n \cap A_n}\} \\ &\quad + \tfrac{1}{2} E_\theta\{(w_n + s_n - \mu)' \nabla^2 g_1(\delta)(w_n + s_n - \mu) 1_{D_n \cap A_n}\} \\ &= R_1 + R_2 + R_3.\end{aligned}$$

It suffices to show that $R_1 + R_2 + R_3 = \mathrm{o}\left(1/\sqrt{n}\right)$. To begin, observe that

$$E_\theta(\|w_n - \mu\|^2) = \mathrm{O}(1/n), \tag{27}$$

$$E_\theta(\|s_n\|^2 1_{D_n}) = \mathrm{O}(1/n) \tag{28}$$

and

$$P_\theta(D_n \cap A_n^c) \leqslant P_\theta(\|w_n - \mu\| > \eta/2) + P_\theta(\|s_n\| 1_{D_n} > \eta/2) = \mathrm{o}(1/n) \tag{29}$$

uniformly in $\theta \in K_1$. In view of part (a), (29) and Schwarz's inequality, we have

$$\begin{aligned} |R_1| &= |E_\xi((Q_n^{ij} - Q_\theta^{ij}) 1_{D_n \cap A_n^c})| \\ &\leqslant \{E_\xi((Q_n^{ij} - Q_\theta^{ij})^2 1_{D_n \cap A_n^c}) P_\xi(D_n \cap A_n^c)\}^{1/2} \\ &= \mathrm{o}(1/\sqrt{n}). \end{aligned}$$

Since $E_\theta(w_n - \mu) = 0$, we can rewrite the second part of R_2 as

$$E_\xi\{(w_n + s_n - \mu) 1_{D_n \cap A_n}\} = -E_\xi\{(w_n - \mu) 1_{D_n^c \cup A_n^c}\} + E_\xi\{s_n 1_{D_n \cap A_n}\}. \tag{30}$$

By Schwarz's inequality, (27), (29) and Lemma 5.6, the first term on the right side of (30) is $\mathrm{o}\left(1/\sqrt{n}\right)$, and write

$$E_\xi\{|s_n| 1_{D_n \cap A_n}\} \leqslant \frac{C}{n} E_\xi[1 + \|y_{0,p}\|^2] = \mathrm{o}\left(\frac{1}{\sqrt{n}}\right).$$

For R_3, since $\nabla^2 g_1(\delta)$ is bounded on $D_n \cap A_n$, we obtain

$$2|R_3| \leqslant C' E_\xi\{\|w_n - \mu\|^2 1_{D_n \cap A_n} + \|s_n\|^2 1_{D_n \cap A_n}\}$$

for some $C' > 0$. So, $R_3 = \mathrm{o}\left(1/\sqrt{n}\right)$ follows from (27) and (28). $\square$

Corollary 5.10. *For any compact convex* $K_0 \subseteq \Omega$,

$$\lim_{n \to \infty} \left[\int_{K_0} E_\theta\{\|Q_n - Q_\theta\|^{2q} 1_{D_n}\} \, \mathrm{d}\theta + \sqrt{n} \int_{K_0} \|E_\theta\{[Q_n - Q_\theta] 1_{D_n}\}\| \, \mathrm{d}\theta \right] = 0.$$

Proof. This follows directly from Proposition 5.9(b), by letting ξ have a slightly larger support than K_0. $\square$

6. Proofs of the main results

The proofs of (6) and (7) are presented in this section. As above ξ denotes a fixed, but arbitrary, twice continuously differentiable density with compact, convex support K in Ω, and $\|\cdot\|$ denotes the spectral norm of a matrix throughout this section.

Proof of (13). Recall that $\mathscr{H}_k^o$ denotes all measurable h for which $|h(z)| \leqslant 1 + \|z\|^k$. The first lemma has the flavor of the proof of (13) with fewer technicalities. Throughout this section, the notation 'essup f' means *essential supremum* of f. $\square$

Lemma 6.11.

$$\lim_{n\to\infty} E_\xi \left[\operatorname*{essup}_{h\in\mathscr{H}_1^o} |E_\xi^n\{h(Z_n)\} - \Phi^p h| 1_{D_n} \right] = 0.$$

Proof. By Stein's identity,

$$E_\xi^n[h(Z_n) - \Phi^p h] = E_\xi^n \left[Uh(Z_n)' \frac{\nabla f_n(Z_n)}{f_n(Z_n)} \right] = E_\xi^n \left\{ Uh(Z_n)' B_n^{-1} \left[\frac{\nabla \xi}{\xi} + \nabla r_n \right] \right\}.$$

Here $\|Uh\|$ is bounded, say $\|Uh\| \leqslant C$ for all $h \in \mathscr{H}_1^o$, and $\|B_n\|^{-1} \leqslant n^{-3/8}$ on D_n. So,

$$\operatorname*{essup}_{h\in\mathscr{H}_1^o} |E_\xi^n\{h(Z_n)\} - \Phi^p h| 1_{D_n} \leqslant Cn^{-3/8} E_\xi^n \left[\left\| \frac{\nabla \xi}{\xi} \right\| + \|\nabla r_n\| \, 1_{D_n} \right],$$

which is independent of h and approaches 0 in the mean. □

With the notation of (13), let

$$R_n(h) = n \left| E_\xi^n[h(Z_n)] - \left\{ \Phi^p h + \frac{1}{\sqrt{n}} (\Phi^p U h)' E_\xi^n \left[Q_\theta \left(\frac{\nabla \xi}{\xi} \right) \right] \right.\right.$$
$$\left.\left. + \frac{1}{n} E_\xi^n \left\{ \operatorname{tr} \left[(\Phi^p V h) Q_\theta \left(\frac{\nabla^2 \xi}{\xi} \right) Q_\theta' \right] \right\} \right\} \right| 1_{D_n}$$

for $h \in \mathscr{H}_2$.

Theorem 6.12. *If ξ is twice continuously differentiable with compact support K, and*

$$\int_\Omega \left\| \frac{\nabla^2 \xi}{\xi} \right\|^\alpha \xi \, d\theta < \infty \tag{31}$$

for some $\alpha > 1$, then

$$\lim_{n\to\infty} E_\xi \left\{ \operatorname*{essup}_{h\in\mathscr{H}_2^o} R_n(h) \right\} = 0. \tag{32}$$

Moreover, (14) *holds for all $h \in \mathscr{H}_2$ and uniformly with respect to $h \in \mathscr{H}_2^o$.*

Proof. Relation (32) is established first. From (13), it suffices to show that

$$\sqrt{n} \left\| \int_{D_n} (\mathrm{I}_n + \mathrm{II}_n) \, dP_\xi \right\| + \int_{D_n} \operatorname*{esssup}_{h\in\mathscr{H}_2^o} (|\mathrm{III}_n(h)| + |\mathrm{IV}_n(h)|) \, dP_\xi \to 0$$

as $n \to \infty$, where I_n–IV_n are defined following (13). These four terms are considered separately. For I_n,

$$\begin{aligned}\left\|\int_{D_n} \sqrt{n}\mathrm{I}_n \,\mathrm{d}P_\xi\right\| &= \sqrt{n}\|E_\xi\{\mathrm{I}_n 1_{D_n}\}\| \\ &= \sqrt{n}\left\|E_\xi\left\{(Q_n - Q_\theta)\left(\frac{\nabla\xi}{\xi}\right)1_{D_n}\right\}\right\| \\ &\leqslant C\sqrt{n}\left\|\int_K E_\theta\{(Q_n - Q_\theta)1_{D_n}\}\nabla\xi(\theta)\,\mathrm{d}\theta\right\| \\ &\leqslant C\sqrt{n}\int_K \|E_\theta[(Q_n - Q_\theta)1_{D_n}]\|\|\nabla\xi(\theta)\|\,\mathrm{d}\theta \\ &\to 0\end{aligned}$$

by Corollary 5.10(b). For II_n, write

$$\sqrt{n}E_\xi[Q_n\nabla r_n(\theta)1_{D_n}] = \sqrt{n}E_\xi[Q_\theta\nabla r_n(\theta)1_{D_n}] + \sqrt{n}E_\xi[(Q_n - Q_\theta)\nabla r_n(\theta)1_{D_n}].$$

The two terms on the right approach zero by Propositions 5.8 and 5.9 and Hölder's inequality.

Next, since Vh is bounded when $h \in \mathscr{H}_2^o$, there is a constant C for which

$$|\mathrm{III}_n(h)|1_{D_n} \leqslant C\|Q_n\|^2\left[2\|\nabla r_n\|\left\|\frac{\nabla\xi}{\xi}\right\| + \|\nabla^2 r_n\| + \|\nabla r_n\|^2\right]1_{D_n},$$

which is independent of h. The expectation of the terms on the right approach zero as $n \to \infty$, by Propositions 5.8 and 5.9. This is clear for $E_\xi[\|Q_n\|^2(\|\nabla^2 r_n\| + \|\nabla r_n\|^2)1_{D_n}]$. For the first term on the right, there is a twice continuously differentiable density $\tilde{\xi}$ with compact convex support and a constant C for which $\|\nabla\xi\| \leqslant C\tilde{\xi}$, as in the proof of Lemma 5.4, and then

$$E_\xi\left\{\|Q_n\|^2\|\nabla r_n\|\left\|\frac{\nabla\xi}{\xi}\right\|1_{D_n}\right\} \leqslant CE_{\tilde{\xi}}[\|Q_n\|^2\|\nabla r_n\|1_{D_n}],$$

which approaches zero as $n \to \infty$, by Propositions 5.8 and 5.9.

For IV_n, first observe that

$$\lim_{n\to\infty} E_\xi\left\|E_\xi^n\left[\frac{\nabla^2\xi}{\xi}\right] - \frac{\nabla^2\xi}{\xi}\right\|^\alpha \to 0 \tag{33}$$

by the Martingale Convergence theorem and (31), and then write

$$\mathrm{IV}_n = \mathrm{IV}_{1,n} + \mathrm{IV}_{2,n} + \mathrm{IV}_{3,n},$$

where

$$\mathrm{IV}_{1,n}(h) = E_\xi^n \operatorname{tr}\left\{Vh(Z_n)Q_n\left[\frac{\nabla^2\xi}{\xi} - E_\xi^n\left(\frac{\nabla^2\xi}{\xi}\right)\right]Q_n'\right\},$$

$$\mathrm{IV}_{2,n}(h) = E_\xi^n \operatorname{tr}\left\{[Vh(Z_n) - \Phi^p Vh]Q_n E_\xi^n\left[\frac{\nabla^2\xi}{\xi}\right]Q_n'\right\}$$

and

$$\mathrm{IV}_{3,n}(h) = E_\xi^n \operatorname{tr}\left\{(\Phi^p Vh)\left[Q_n\left(\frac{\nabla^2\xi}{\xi}\right)Q_n' - Q_\theta\left(\frac{\nabla^2\xi}{\xi}\right)Q_\theta'\right]\right\}.$$

As in the analysis of $\mathrm{III}_n(h)$, there is a constant C for which

$$|\mathrm{IV}_{3,n}(h)| \leqslant C(\|Q_n - Q_\theta\|)(\|Q_n + Q_\theta\|)\left\|\frac{\nabla^2\xi}{\xi}\right\|$$

which is independent of h and approaches zero in the mean by Proposition 5.9 and (33). For $\mathrm{IV}_{1,n}$ and $\mathrm{IV}_{2,n}$

$$E_\xi\left[\operatorname*{essup}_{h\in\mathscr{H}_2^o}|\mathrm{IV}_{1,n}(h)|1_{D_n}\right] \leqslant C E_\xi\left\{\left\|E_\xi^n\left[\frac{\nabla^2\xi}{\xi}\right] - \frac{\nabla^2\xi}{\xi}\right\|\|Q_n\|^2\right\} \to 0$$

as $n \to \infty$, by Proposition 5.9, (33), and Hölder's inequality. Similarly, since Vh is bounded when $h \in \mathscr{H}_2^o$

$$E_\xi\left[\operatorname*{essup}_{h\in\mathscr{H}_2^o}|\mathrm{IV}_{2,n}(h)|1_{D_n}\right]$$

$$\leqslant E_\xi\left\{\operatorname*{esssup}_{h\in\mathscr{H}_2^o}\|E_\xi^n[Vh(Z_n) - \Phi^p Vh]\|\left\|E_\xi^n\left[\frac{\nabla^2\xi}{\xi}\right]\right\|\|Q_n\|^2 1_{D_n}\right\} \to 0$$

by Proposition 5.9, Lemma 6.11, (33), and Hölder's inequality. This completes the proof of (32).

For (14), let $\bar{R}_n = \operatorname{essup}_{h\in\mathscr{H}_2^o} R_n(h)$. Then there is a constant C for which the difference between $E_\xi[h(Z_n)]$ and its approximation in (14) is at most

$$\int_{D_n}\bar{R}_n\,\mathrm{d}P_\xi + C\int_{D_n^c}\left[1 + \|Z_n\|^2 + \frac{1}{\sqrt{n}}\|Q_\theta\|\left\|\frac{\nabla\xi}{\xi}\right\| + \frac{1}{n}\|Q_\theta\|^2\left\|\frac{\nabla^2\xi}{\xi}\right\|\right]P_\xi.$$

The first term here is $\mathrm{o}(1/n)$ by the Theorem, the second by Lemma 5.7, and the remaining two by Lemma 5.6. □

Studentization: Two more auxiliary results are needed for the transition from Z_n to Z_n^* and T_n^*. As in the last section, let K_1 and K_2 be compact convex sets for which $K \subset K_1^o \subset K_1 \subset K_2^o \subset K_2 \subset \Omega$.

Proposition 6.13. *Let D_n satisfy* (23). *If g is continuous on Ω, then*

$$\lim_{n\to\infty}\int_{D_n}|g(\hat\theta_n)-g(\theta)|\,\mathrm{d}P_\xi=0$$

and if g is twice continuously differentiable on Ω, then

$$\int_{D_n}[g(\theta)-g(\hat\theta_n)]\,\mathrm{d}P_\xi=\mathrm{o}\left(\frac{1}{\sqrt{n}}\right).$$

Proof. By compactness and continuity, g and its derivatives (if continuous) are bounded on K_2. The first assertion then follows directly from the Dominated convergence theorem. For the second, there is a constant C for which

$$|g(\theta)-g(\hat\theta_n)-\nabla g(\hat\theta_n)'(\theta-\hat\theta_n)|\leqslant C\|\theta-\hat\theta_n\|^2$$

for all $\theta\in K$ on D_n; and $E_\xi(\|\theta-\hat\theta_n\|^2 1_{D_n})=\mathrm{o}\left(1/\sqrt{n}\right)$ by Lemma 5.6. Further, using $E^n_\xi(\theta-\hat\theta_n)=B_n'^{-1}E^n_\xi(Z_n)$ and Proposition 2.1, there is a C for which

$$\begin{aligned}\left|\int_{D_n}\nabla g(\hat\theta_n)'(\theta-\hat\theta_n)\,\mathrm{d}P_\xi\right| &\leqslant C\int_{D_n}\|E^n_\xi(\theta-\hat\theta_n)\|\,\mathrm{d}P_\xi\\ &= C\int_{D_n}\left\|(X_n'X_n)^{-1}E^n_\xi\left[\frac{\nabla\xi}{\xi}+\nabla r_n(\theta)\right]\right\|\mathrm{d}P_\xi\\ &\leqslant Cn^{-3/4}\int_{D_n}\left[\left\|\frac{\nabla\xi}{\xi}\right\|+\|r_n(\theta)\|\right]\mathrm{d}P_\xi,\end{aligned}$$

which is $\mathrm{o}\left(1/\sqrt{n}\right)$ by Proposition 5.8. □

Given a function $h\in\mathscr{H}$, a $t>0$, a $v\in\mathbb{R}^p$, and a $p\times p$ non-singular matrix Γ, let

$$h^*(z)=h[t^{-1/2}\Gamma^{-1}(z-v)],$$

$$\begin{aligned}\Psi_0(h;v,t,\Gamma)= & -(\Phi^pUh)'v+\mathrm{tr}\{(\Phi^pVh)[vv'-(\Gamma+\Gamma'-2I_p)]\}\\ & -\mathrm{tr}(\Phi^pVh)(t-1)-(\Phi^p_3h)'v(t-1)+\tfrac{1}{2}(\Phi^p_4h)(t-1)^2,\end{aligned}$$

$$\Psi_1(h;v,t,\Gamma)=-2(\Phi^pVh)v+(\Phi^p_3h)(t-1),$$

where

$$\Phi^p_3h=\frac{1}{2}\int_{\Re^p}[p+1-\|z\|^2]zh(z)\Phi^p\{\mathrm{d}z\},$$

$$\Phi^p_4h=\int_{\Re^p}\left\{\frac{1}{4}[\|z\|^2-p]^2-\frac{1}{2}p\right\}h(z)\Phi^p\{\mathrm{d}z\}.$$

Lemma 6.14. *There is a constant C for which*

$$|\Phi^p h^* - \Phi^p h - \Psi_0(h; v, t, \Gamma)| \leqslant C[\|v\|^3 + |t-1|^3 + \|\Gamma - I_p\|^{3/2}],$$

$$\|\Phi^p U h^* - \Phi^p U h - \Psi_1(h; v, t, \Gamma)\| \leqslant C[\|v\|^2 + |t-1|^2 + \|\Gamma - I_p\|]$$

and

$$\|\Phi^p V h^* - \Phi^p V h\| \leqslant C[\|v\| + |t-1| + \|\Gamma - I_p\|]$$

for all $\|v\| \leqslant 1$, $\frac{1}{2} \leqslant t \leqslant \frac{3}{2}$, $\|\Gamma - I_p\| \leqslant \frac{1}{2}$, *and* $h \in \mathscr{H}_0$.

Proof. As in Woodroofe and Coad (1997),

$$\Phi^p h^* = \int_{\Re^p} h(x)\varphi(x; v, t, \Gamma)\,\mathrm{d}x,$$

where

$$\varphi(x; v, t, \Gamma) = t^{1/2p}|\det(\Gamma)|\varphi\left(\sqrt{t}\Gamma x + v\right).$$

The latter forms an exponential family of densities, so that the integral is infinitely differentiable. The first assertion then follows from a Taylor series expansion and the identity $\Gamma\Gamma' - I_p = (\Gamma + \Gamma' - 2I_p) + (\Gamma - I_p)(\Gamma' - I_p)$. The others may be established similarly. □

Recall that $Z_n^* = \hat{\Gamma}_n^{-1}(Z_n - \hat{\mu}_n)$ and $T_n^* = \tilde{\sigma}_n^{-1}\hat{\Gamma}_n^{-1}(Z_n - \tilde{\sigma}_n\hat{\mu}_n)$.

Theorem 6.15. *Let* $\hat{\mu}_n$ *and* $\hat{\Gamma}_n$ *be as in* (18) *and* (20). *If* ξ *is a twice continuously differentiable density with compact support for which* (31) *holds, then* (6) *and* (7) *hold. In fact,*

$$\left|\int_\Omega [E_\theta[h(Z_n^*)] - \Phi^p h]\xi(\theta)\,\mathrm{d}\theta\right| = \mathrm{o}\left(\frac{1}{n}\right) \tag{34}$$

and

$$\left|\int_\Omega \left[E_\theta[h(T_n^*)] - \Phi^p h - \frac{1}{n}\Phi_4^p h\right]\xi(\theta)\,\mathrm{d}\theta\right| = \mathrm{o}\left(\frac{1}{n}\right) \tag{35}$$

for all $h \in \mathscr{H}_0$ *and uniformly with respect to* $h \in \mathscr{H}_0^o$.

Proof. Relations (6) and (7) follow from (34) and (35) by letting h be an indicator function and using the relation $G_n^p h = \Phi^p h + \Phi_4^p h/n + \mathrm{o}(1/n)$ for $h \in \mathscr{H}$. Only the proof of (35) is given; that of (34) is similar and simpler. Let D_n^o be the right side of (23),

$$D_n = D_n^o \cap \left\{\tfrac{1}{2} \leqslant \tilde{\sigma}_n^2 \leqslant \tfrac{3}{2}\right\}.$$

Then D_n satisfies (23) and (24). If $h \in \mathscr{H}_0^o$, let

$$h_n(z) = h[\tilde{\sigma}_n^{-1}\hat{\Gamma}_n^{-1}(z - \tilde{\sigma}_n\hat{\mu}_n)].$$

Then

$$E_\xi[h(T_n^*)] - \Phi^p h = E_\xi\{E_\xi^n[h(T_n^*) - \Phi^p h]\mathbf{1}_{D_n}\} + \mathrm{o}\left(\frac{1}{n}\right),$$

$$E_\xi^n[h(T_n^*)] = E_\xi^n[h_n(Z_n)]$$

and

$$E_\xi^n[h_n(Z_n)] - \Phi^p h = \{E_\xi^n[h_n(Z_n)] - \Phi^p h_n\} + [\Phi^p h_n - \Phi^p h].$$

On D_n,

$$E_\xi^n[h_n(Z_n)] - \Phi^p h_n = \frac{1}{\sqrt{n}}(\Phi^p U h_n)' E_\xi^n\left[Q_\theta\left(\frac{\nabla\xi}{\xi}\right)\right] + \frac{1}{n} E_\xi^n \operatorname{tr}\left\{(\Phi^p V h_n) Q_\theta\left(\frac{\nabla\xi}{\xi}\right) Q_\theta'\right\} + \frac{1}{n} R_n(h_n).$$

Now $\Phi^p h_n$, $\Phi^p U h_n$, and $\Phi^p V h_n$ may be approximated using Lemma 6.14. After this approximation and some algebra,

$$E_\xi^n[h_n(Z_n)] - \Phi^p h = \frac{1}{\sqrt{n}}(\Phi^p U h)' \mathrm{I}_n^* + \frac{1}{n}\operatorname{tr}\{(\Phi^p V h)[\mathrm{II}_n^* - n(\tilde{\sigma}_n^2 - 1)]\} - (\Phi_3^p h)' \tilde{\sigma}_n \hat{\mu}_n (\tilde{\sigma}_n^2 - 1) + \frac{1}{2n}(\Phi_4^p h) n (\tilde{\sigma}_n^2 - 1)^2 + \frac{1}{n}\tilde{R}_n(h),$$

where

$$\mathrm{I}_n^* = E_\xi^n\left[Q_\theta\left(\frac{\nabla\xi}{\xi}\right)\right] - \sqrt{n}\tilde{\sigma}_n\hat{\mu}_n,$$

$$\mathrm{II}_n^* = E_\xi^n\left[Q_\theta\left(\frac{\nabla^2\xi}{\xi}\right) Q_\theta' + n\tilde{\sigma}_n^2\hat{\mu}_n\hat{\mu}_n' - \hat{\Delta}_n - 2Q_\theta\frac{\nabla\xi}{\xi}\sqrt{n}\tilde{\sigma}_n\hat{\mu}_n'\right]$$

and $\lim_{n\to\infty} E_\xi[\operatorname{essup}_{h\in\mathscr{H}_2^o}|\tilde{R}_n(h)|] = 0$. Here $\lim_{n\to\infty} E_\xi[n(\tilde{\sigma}_n^2 - 1)^2\mathbf{1}_{D_n}] = 2$, by Lemma 5.4 and the preceding remark. So, it suffices to show that

$$E_\xi[(\tilde{\sigma}_n^2 - 1)\mathbf{1}_{D_n}] = \mathrm{o}(1/n),$$

$$E_\xi[\tilde{\sigma}_n\hat{\mu}_n(\tilde{\sigma}_n^2 - 1)\mathbf{1}_{D_n}] = \mathrm{o}(1/n),$$

$$E_\xi(\mathrm{I}_n^*\mathbf{1}_{D_n}) = \mathrm{o}\left(1/\sqrt{n}\right),$$

$$E_\xi(\mathrm{II}_n^* D_n) \to 0.$$

The first of these is clear, since

$$E_\xi[(\tilde{\sigma}_n^2 - 1)\mathbf{1}_{D_n}] = -E_\xi[(\tilde{\sigma}_n^2 - 1)\mathbf{1}_{D_n^c}] + \mathrm{o}(1/n) = \mathrm{o}(1/n),$$

by Lemma 5.4 and Schwarz's inequality, and the second follows since $n\tilde{\sigma}_n\hat{\mu}_n(\tilde{\sigma}_n^2 - 1)$ has a limiting distribution with mean 0. For II_n^*, it is easily seen that $\mathrm{II}_n^* \to^p M(\theta)$, where

$$M(\theta) = Q_\theta\left(\frac{\nabla^2\xi}{\xi}\right)Q'_\theta + (Q_\theta^\# \mathbf{1})(Q_\theta^\# \mathbf{1})' - \Delta(\theta) - 2Q_\theta\frac{\nabla\xi}{\xi}(Q_\theta^\# \mathbf{1})'$$

and that

$$E_\xi[\mathrm{II}_n^* 1_{D_n}] \to \int_\Omega M(\theta)\xi(\theta)\,\mathrm{d}\theta = 0,$$

where the final equality follows from an integration by parts. See Lemma 2 (Woodroofe and Coad, 1997).

The term involving I_n^* is more delicate. Here

$$\begin{aligned}
\mathrm{I}_n^* &= E_\xi^n\left[Q_\theta\left(\frac{\nabla\xi}{\xi}\right) + Q_\theta^\# \mathbf{1}\right]\mathbf{1}_{D_n} + E_\xi^n[\tilde{\sigma}_n Q_{\hat{\theta}_n}^\# \mathbf{1} - Q_\theta^\# \mathbf{1}]\mathbf{1}_{D_n} \\
&= \mathrm{I}_{1,n}^* + \mathrm{I}_{2,n}^*, \\
E_\xi(\mathrm{I}_{1,n}^*) &= \int_\Omega \left[Q_\theta\left(\frac{\nabla\xi}{\xi}\right) + Q_\theta^\# \mathbf{1}\right] P_\theta(D_n)\xi(\theta)\,\mathrm{d}\theta \\
&= -\int_\Omega \left[Q_\theta\left(\frac{\nabla\xi}{\xi}\right) + Q_\theta^\# \mathbf{1}\right] P_\theta(D_n^c)\xi(\theta)\,\mathrm{d}\theta \\
&= \mathrm{o}(1/n);
\end{aligned}$$

and

$$E_\xi(\mathrm{I}_{2,n}^*) = E_\xi\left\{\left[\tilde{\sigma}_n Q_{\hat{\theta}_n}^\# \mathbf{1} - Q_\theta^\# \mathbf{1}\right]\mathbf{1}_{D_n}\right\} = \mathrm{o}\left(1/\sqrt{n}\right)$$

by Lemmas 5.4 and 5.6 and Proposition 6.13. This completes the proof of (35). The proof of (34) is similar, but the terms involving $\tilde{\sigma}_n^2$ are absent. □

7. Discussions

We have derived approximate confidence intervals for a stationary, Gaussian autoregressive process of order p. Simulation experiments for AR(2) processes with sample sizes $n = 10, 20, 50$ show excellent agreement with the theoretical results, recalling that they predict better agreement in the symmetric case than for the one-sided one.

Though this paper mainly concerns setting confidence intervals from a frequentist point of view, the integrable expansion derived here has a close connection with Bayesian models. In the Bayesian literature, there has been an extensive study on developing priors that match asymptotically the coverage probabilities of Bayesian credible sets with the corresponding frequentist probabilities. Such priors are referred to as 'probability matching priors.' See, for example, Ghosh (1994), Datta and Mukerjee (2004). Our work suggests an equation for the matching prior. Letting $\rho = \log\xi$ and approximating the coefficient of $1/\sqrt{n}$ in (13) suggests that the frequentist approximation (16) and posterior expansion (13) will agree to order $\mathrm{o}\left(1/\sqrt{n}\right)$ if

$$Q_\theta \nabla \rho(\theta) = -Q_\theta^\# \mathbf{1}.$$

This has the solution

$$\rho(\theta) = \rho(0) - \int_0^1 \theta' Q_{t\theta}^{-1} Q_{t\theta}^\# \mathbf{1}\, \mathrm{d}t$$

or

$$\xi(\theta) \propto \exp\left[-\int_0^1 \theta' Q_{t\theta}^{-1} Q_{t\theta}^\# \mathbf{1}\, \mathrm{d}t\right]. \tag{36}$$

Of course, our conditions are not satisfied by (36), so that (36) can be at most suggestive, but it is that at least. If $p=1$, then the unique solution to (36) is the Jeffreys prior, $\xi(\theta) \propto \sqrt{\mathscr{I}_\theta}$, where $\mathscr{I}_\theta$ is the information, but if $p \geqslant 2$, then (36) need not be the Jeffreys prior. For $p=2$, it is (after straightforward, but lengthy calculations)

$$\xi(\theta) \propto \sqrt{\frac{1+\theta_2}{1-\theta_2}} \sqrt{|\mathscr{I}_\theta|},$$

where now $|\mathscr{I}_\theta|$ is the determinant of the information matrix.

Further investigation of (36), including more general conditions on the prior for its validity, is one open problem for future research. Another is to extend the result to ARMA processes. A natural approach is to write an ARMA process as an AR process of infinite order, where the sequence of coefficients in the latter depends on finitely many parameters, but this may be messy (at best), because the likelihood function for an ARMA process is complicated.

Acknowledgements

The authors would like to thank the referees and associate editor for their comments on this paper. Both authors are supported by the National Science Foundation, and the first author is also partially supported by National Science Council of Taiwan.

References

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods. Springer, New York.

Coad, D.S., Woodroofe, M., 1998. Approximate bias calculations for sequentially designed experiments. Sequential Anal. 17, 1–31.

Datta, G.S., Mukerjee, R., 2004. Probability Matching Priors: Higher Order Asymptotics. Series: Lecture Notes in Statistics, vol. 178. Springer, New York.

Ghosh, J.K., 1994. Higher Order Asymptotics. IMS, Hayward, CA.

Mann, H.B., Wald, A., 1943a. On stochastic limit and order relations. Ann. Math. Statist. 14, 217–226.

Mann, H.B., Wald, A., 1943b. On the statistical treatment of linear stochastic difference equations. Econometrica 11, 173–220.

Stein, C., 1981. Estimation of the mean of a multivariate normal distribution. Ann. Statist. 9, 1135–1151.
Stein, C., 1987. Approximate Computation of Expectations. IMS, Hayward, CA.
Tanaka, K., 1984. An asymptotic expansions associated with the maximum likelihood estimators in ARMA models. J. R. Statist. Soc. B 46, 58–67.
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., 1992. Integrable expansions for posterior distributions for one-parameter exponential families. Statist. Sinica 2, 91–111.
Woodroofe, M., Coad, D.S., 1997. Corrected confidence sets for sequentially designed experiments. Statist. Sinica 7, 53–74.
Woodroofe, M., Coad, D.S., 2002. Corrected confidence sets for sequentially designed experiments: examples. Sequential Anal. 21, 191–218.