L_1-Statistical Procedures and Related Topics
IMS Lecture Notes – Monograph Series (1997) Volume 31

Some limit theory for L_1-estimators in autoregressive models under general conditions

Keith Knight

University of Toronto, Ontario, Canada

Abstract: It is well-known that L_1-estimators of autogressive parameters are asymptotically Normal if the distribution function of the errors, $F(x)$, has $F'(0) = \lambda > 0$. In this paper, we derive limiting distributions of L_1-estimators under more general assumptions on F. Second-order representations are also derived.

Key words: Autoregressive processes, limiting destributions, Bahadur-Kiefer theorems.

AMS subject classification: 62F12, 62M10, 60F05.

1 Introduction

L_1 estimation provides a somewhat robust alternative to least squares estimation for autoregressive models. Define a p-th order autoregressive (AR(p)) process

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t \tag{1}$$

where $\{\varepsilon_t\}$ are independent, identically distributed (i.i.d.) random variables such that (a) $E(\varepsilon_t^2) < \infty$; (b) ε_t has median 0; (c) $F(x) = P(\varepsilon_t \le x)$ is continuous at $x = 0$.

We will assume that the process $\{Y_t\}$ is stationary; for this, we require that

$$\sum_{k=1}^{p} \phi_k z^k \neq 1$$

for all complex z with modulus $|z| \le 1$. Throughout this paper, we will assume the model (1) with the intercept ϕ_0; however, all of the results given in the paper will go through (with appropriate modifications) if ϕ_0 is suppressed and only $\phi_1, \cdots, \phi_p$ are estimated.

Least squares (or some related method) is typically used to estimate the parameters in the model (1). However, when the ε_t's have heavy tails, least squares is inefficient compared to some other methods; one such method is L_1-estimation. We define L_1-estimators, $\widehat{\phi}_0, \widehat{\phi}_1, \cdots, \widehat{\phi}_p$, to minimize the objective function

$$g(v_0, v_1, \cdots, v_p) = \sum_{t=1}^{n} |Y_t - v_0 - v_1 Y_{t-1} - \cdots - v_p Y_{t-p}|. \tag{2}$$

(This assumes that we have $n+p$ observations but asymptotically has no effect.) It is well-known (see Pollard, 1991; Wang and Wang, 1996) that the asymptotic behaviour of L_1 estimators depends on the behaviour of the distribution function $F(x)$ for x close to 0. For example, if $F'(0) = \lambda > 0$ then we have

$$\sqrt{n}(\widehat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}) \to_d N_{p+1}(\mathbf{0}, C/(4\lambda^2)) \quad \text{as } n \to \infty$$

where C is a $(p+1) \times (p+1)$ matrix defined to be

$$C = E[\boldsymbol{X}_t \boldsymbol{X}_t^T] \tag{3}$$

where $\boldsymbol{X}_t = (1, Y_{t-1}, \cdots, Y_{t-p})^T$. Note that, contrary to popular belief, it is not necessary for F to be absolutely continuous to have asymptotic normality.

The assumption that $F'(0) = \lambda > 0$ is quite strong in the sense that it is difficult to verify; given even very large samples, it is difficult to distinguish between a density which is finite at 0 and one which has a singularity at 0. For the sample median (which is the L_1-estimator of location), it has been shown that (for example, by de Haan and Taconis-Haantjes, 1979) that the rate of convergence depends on the behaviour of $F(x)$ for x close to the population median; see also Smirnov (1952) who derives the domains of attraction for sample quantiles. Similarly, it is not necessary for F to be differentiable in order to find a limiting distribution for the L_1-estimator $\widehat{\boldsymbol{\phi}}_n$. We will assume that for some sequence $\{a_n\}$ with $a_n \to \infty$, there exists a strictly increasing function ψ such that

$$\sqrt{n}(F(t/a_n) - F(0)) = \psi(t) + r_n(t) \tag{4}$$

where $r_n(t) \to 0$ as $n \to \infty$ for each t. Also define

$$\Psi(t) = \int_0^t \psi(s)\, ds \tag{5}$$

and

$$R_n(t) = \int_0^t r_n(s)\,ds. \tag{6}$$

Note that since $\psi(t)$ is strictly increasing, $\Psi(t)$ will be strictly convex.

The formulation given above provides a great deal of flexibility. For example, suppose that

$$F(x) - F(0) = \lambda \mathrm{sgn}(x)|x|^\alpha L(|x|)$$

for x in a neighbourhood of 0 where $\alpha > 0$ and L is a slowly varying function at 0. ($\mathrm{sgn}(x) = 1$ if x is positive and -1 if x is negative.) In this case, we can take

$$a_n = n^{1/(2\alpha)} L^*(n) \quad \text{and} \quad \psi(t) = \lambda \mathrm{sgn}(t)|t|^\alpha$$

where L^* is a slowly varying function at infinity. When $F(x)$ is differentiable at $x = 0$ with $F'(0) = \lambda > 0$ then $\psi(t) = \lambda t$ and $a_n = \sqrt{n}$; however, if $F(x) - F(0) = \lambda x \ln(|x|^{-1})$ for x close to 0 then $\psi(t) = \lambda t$ with $a_n = \sqrt{n}\ln(n)/2$. If $F(x)$ is not differentiable at $x = 0$ but has positive one-sided derivatives λ^+ and λ^- then

$$\psi(t) = \begin{cases} \lambda^+ t & \text{for } t \geq 0 \\ \lambda^- t & \text{for } t < 0 \end{cases}.$$

(This occurs, for example, if the density has a jump at 0.)

In Section 2, we will determine the limiting distribution of the L_1-estimator under the general conditions on F described above, we will define

$$Z_n(\boldsymbol{u}) = \frac{a_n}{\sqrt{n}} \sum_{t=1}^{n} \left[|\varepsilon_t - \boldsymbol{u}^T \boldsymbol{X}_t / a_n| - |\varepsilon_t| \right]. \tag{7}$$

Note that Z_n is minimized at $\boldsymbol{u} = a_n(\widehat{\boldsymbol{\phi}}_n - \boldsymbol{\phi})$. Z_n is a convex function and hence if the finite dimensional distributions converge weakly to those of a convex function Z, it follows that

$$a_n(\widehat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}) \to_d \mathrm{argmin}(Z)$$

provided $\mathrm{argmin}(Z)$ is almost surely unique (Geyer, 1996). What is interesting is that only finite dimensional weak convergence is needed and not any sort of functional weak convergence (although this is implied by the finite dimensional convergence for convex functions).

In Section 3, we will obtain an "in distribution" Bahadur-Kiefer representation for the L_1-estimator under the general conditions described above. This will be done by approximating Z_n by an appropriate function Z_n^* and looking at the limiting behaviour of $n^{1/4}(Z_n - Z_n^*)$.

2 Limiting distributions

In order to derive the limiting distributions of the L_1 estimators, we will assume the following regularity conditions.

(A1) $\{\varepsilon_i\}$ are 0 median, finite variance i.i.d. random variables with distribution function F satisfying (4) for some $\psi(t)$ and $r_n(t)$.

(A2) For each $\boldsymbol{u}$,

$$E[\Psi(\boldsymbol{u}^T\boldsymbol{X}_t)] = \tau(\boldsymbol{u}) < \infty$$

where $\Psi(t)$ is defined in (5) and $\tau(\boldsymbol{u})$ is a strictly convex function.

(A3) For each $\boldsymbol{u}$,

$$\frac{1}{n}\sum_{i=1}^{n} R_n(\boldsymbol{u}^T\boldsymbol{X}_t) \to_p 0$$

as $n \to \infty$ where $R_n(t)$ is defined in (6).

Note that condition (A1) implies that $E[(\boldsymbol{u}^T\boldsymbol{X}_t)^2] < \infty$; thus, depending on the exact form of Ψ, condition (A2) may be implied by (A1). A sufficient condition for (A3) is $E[|R_n(\boldsymbol{u}^T\boldsymbol{X}_t)|] \to 0$.

Theorem 1 *Suppose that $\{Y_t\}$ is an AR(p) process satisfying (1) and that $Z_n(\boldsymbol{u})$ is as defined in (7). Then under conditions (A1), (A2) and (A3),*

$$(Z_n(\boldsymbol{u}_1), \cdots, Z_n(\boldsymbol{u}_k)) \to_d (Z(\boldsymbol{u}_1), \cdots, Z(\boldsymbol{u}_k))$$

as $n \to \infty$ where

$$Z(\boldsymbol{u}) = \boldsymbol{u}^T\boldsymbol{W} + 2\tau(\boldsymbol{u})$$

with $\boldsymbol{W}$ a (p+1)-variate Normal random vector with mean $\boldsymbol{0}$ and covariance matrix C defined in (3).

Proof: We will use the identity

$$|x-y| - |x| = y[I(x<0) - I(x>0)] + 2\int_0^y [I(x \le s) - I(x<0)]\, ds$$

which is valid for $x \neq 0$. ($I(A)$ is the indicator function of the set A.) Now

$$Z_n(\boldsymbol{u}) = Z_n^{(1)}(\boldsymbol{u}) + Z_n^{(2)}(\boldsymbol{u})$$

where

$$Z_n^{(1)}(\boldsymbol{u}) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{X}_t^T\boldsymbol{u}[I(\varepsilon_t < 0) - I(\varepsilon_t > 0)]$$

$$\text{and} \quad Z_n^{(2)}(\boldsymbol{u}) = \frac{2a_n}{\sqrt{n}} \sum_{t=1}^{n} \int_0^{v_{nt}} [I(\varepsilon_t \le s) - I(\varepsilon_t \le 0)]\, ds = \sum_{t=1}^{n} Z_{nt}^{(2)}(\boldsymbol{u})$$

(with $v_{nt} = \boldsymbol{X}_t^T \boldsymbol{u}/a_n$). Since, for each $\boldsymbol{u}$, the summands in $Z_n^{(1)}(\boldsymbol{u})$ are stationary martingale differences with finite variance, it follows from a martingale central limit theorem that

$$Z_n^{(1)}(\boldsymbol{u}) \to_d \boldsymbol{u}^T \boldsymbol{W}$$

and the convergence in distribution holds for any finite collection of $\boldsymbol{u}$'s. For $Z_n^{(2)}(\boldsymbol{u})$, we have

$$Z_n^{(2)}(\boldsymbol{u}) = \sum_{t=1}^{n} E(Z_{nt}^{(2)}(\boldsymbol{u})) + \sum_{t=1}^{n} (Z_{nt}^{(2)}(\boldsymbol{u}) - E(Z_{nt}^{(2)}(\boldsymbol{u}))).$$

Letting $v_t = \boldsymbol{X}_t^T \boldsymbol{u} = a_n v_{nt}$, it follows that

$$\begin{aligned} \sum_{t=1}^{n} E(Z_{nt}^{(2)}(\boldsymbol{u})) &= \frac{2a_n}{\sqrt{n}} \sum_{t=1}^{n} \int_0^{v_{nt}} (F(s) - F(0))\, ds \\ &= \frac{2}{n} \sum_{t=1}^{n} \int_0^{v_t} \sqrt{n}(F(s/a_n) - F(0))\, ds \\ &= \frac{2}{n} \sum_{t=1}^{n} \left[\Psi(\boldsymbol{u}^T \boldsymbol{X}_t) + R_n(\boldsymbol{u}^T \boldsymbol{X}_t) \right] \\ &\to 2\tau(\boldsymbol{u}) \end{aligned}$$

where R_n is defined in (6). For the remainder term in $Z_n^{(2)}(\boldsymbol{u})$, we have (since the summands are again martingale differences)

$$\begin{aligned} \operatorname{Var}(Z_n^{(2)}(\boldsymbol{u})) &= \sum_{t=1}^{n} E[(Z_{nt}^{(2)}(\boldsymbol{u}) - E(Z_{nt}^{(2)}(\boldsymbol{u})))^2] \\ &\le \frac{2}{\sqrt{n}} \max_{1 \le t \le n} |\boldsymbol{X}_t^T \boldsymbol{u}| \sum_{t=1}^{n} E(Z_{nt}^{(2)}(\boldsymbol{u})) \\ &= \frac{2}{\sqrt{n}} \max_{1 \le t \le n} |\boldsymbol{X}_t^T \boldsymbol{u}| E(Z_n^{(2)}(\boldsymbol{u})). \end{aligned}$$

$\{\boldsymbol{X}_t^T \boldsymbol{u}\}$ is stationary with finite second moment and so

$$\max_{1 \le t \le n} |\boldsymbol{X}_t^T \boldsymbol{u}| \to_p 0.$$

Thus

$$Z_n^{(2)}(\boldsymbol{u}) - E(Z_n^{(2)}(\boldsymbol{u})) \to_p 0 \quad \text{as } n \to \infty$$

and so $Z_n^{(2)}(\boldsymbol{u}) \to_p 2\tau(\boldsymbol{u})$. Thus we have

$$Z_n(\boldsymbol{u}) \to_d \boldsymbol{u}^T \boldsymbol{W} + 2\tau(\boldsymbol{u}) = Z(\boldsymbol{u})$$

and the finite dimensional convergence holds trivially. □

The following corollary gives us a representation of the limiting distribution of $a_n(\widehat{\boldsymbol{\phi}}_n - \boldsymbol{\phi})$.

Corollary 2 *Let $\widehat{\boldsymbol{\phi}}_n$ minimize (2). Under the assumptions of Theorem 1,*

$$a_n(\widehat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}) \to_d \mathit{argmin}(Z)$$

as $n \to \infty$.

Proof: Since τ is strictly convex, Z_α is strictly convex and so has a unique minimum. The result follows from Geyer (1996). □

The limiting distribution given in Corollary 2 will not be normal unless the function $\tau(\boldsymbol{u})$ is quadratic. In the following example, we illustrate the computation of the limiting distribution in a special case.

Example 1 Consider the AR(1) process

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t$$

where the ε_t's are i.i.d. random variables with density

$$f_\alpha(x) = \frac{|x|^{\alpha-1}\exp(-|x|)}{2\Gamma(\alpha)}$$

for some $\alpha > 0$. (This is a two-sided Gamma distribution.) For x close to 0, we have

$$F_\alpha(x) - F_\alpha(0) \approx \frac{\operatorname{sgn}(x)|x|^\alpha \exp(-x)}{2\Gamma(\alpha+1)}$$

and so setting $a_n = n^{1/(2\alpha)}$, we get

$$\sqrt{n}(F_\alpha(t/a_n) - F_\alpha(0)) \to \psi_\alpha(t) = \frac{1}{2\Gamma(\alpha+1)}\operatorname{sgn}(t)|t|^\alpha$$

with

$$|\sqrt{n}(F_\alpha(t/a_n) - F_\alpha(0)) - \psi_\alpha(t)| \le k(\alpha)\frac{|t|^{\alpha+1}}{n^{1/(2\alpha)}}.$$

It is easy to verify that conditions (A1), (A2) and (A3) are all satisfied (since $E[|Y_t|^r] < \infty$ for all $r > 0$) and so for given ϕ_0, ϕ_1, the limiting objective function in Theorem 1 is

$$Z_\alpha(u_0, u_1) = u_0 W_0 + u_1 W_1 + \frac{1}{\Gamma(\alpha+2)} E[|u_0 + u_1 Y_1|^{\alpha+1}]$$

where W_0 and W_1 are zero mean Normal random variables with $\mathrm{Var}(W_0) = 1$, $\mathrm{Var}(W_1) = E(Y_1^2)$ and $\mathrm{Cov}(W_0, W_1) = E(Y_1)$. By differentiation, we determine the minimizers of Z_α, $\widehat{U}_0$ and $\widehat{U}_1$, to satisfy the equations

$$\begin{aligned} W_0 + \frac{1}{\Gamma(\alpha+1)} d_0(\widehat{U}_0, \widehat{U}_1) &= 0 \\ W_1 + \frac{1}{\Gamma(\alpha+1)} d_1(\widehat{U}_0, \widehat{U}_1) &= 0 \end{aligned}$$

where

$$\begin{aligned} d_0(u_0, u_1) &= E\left[\mathrm{sgn}(u_0 + u_1 Y_1)|u_0 + u_1 Y_1|^\alpha\right] \quad \text{and} \\ d_1(u_0, u_1) &= E\left[\mathrm{sgn}(u_0 + u_1 Y_1) Y_1 |u_0 + u_1 Y_1|^\alpha\right]. \end{aligned}$$

($\mathrm{sgn}(x) = 1$ or -1 depending on whether x is positive or negative.)

If $f_W(w_0, w_1)$ is the joint density of (W_0, W_1), it then follows that the joint density of $(\widehat{U}_0, \widehat{U}_1)$ (that is, the limiting density of $a_n(\widehat{\boldsymbol{\phi}}_n - \boldsymbol{\phi})$) is

$$f_U(u_0, u_1) = \frac{1}{\Gamma(\alpha)^2} f_W\left(\frac{d_0(u_0, u_1)}{\Gamma(\alpha+1)}, \frac{d_1(u_0, u_1)}{\Gamma(\alpha+1)}\right)\left(d_{00}(u_0, u_1) d_{11}(u_0, u_1) - d_{10}^2(u_0, u_1)\right)$$

where

$$\begin{aligned} d_{00}(u_0, u_1) &= E\left[|u_0 + u_1 Y_1|^{\alpha-1}\right] \\ d_{11}(u_0, u_1) &= E\left[Y_1^2 |u_0 + u_1 Y_1|^{\alpha-1}\right] \quad \text{and} \\ d_{10}(u_0, u_1) &= E\left[Y_1 |u_0 + u_1 Y_1|^{\alpha-1}\right]. \end{aligned}$$

The density f_U cannot easily be computed analytically (unless $\phi_1 = 0$) but can be computed feasibly using Monte Carlo sampling.

3 Second order properties

It follows from the proof of Theorem 1 that we can approximate Z_n by the function

$$Z_n^*(\boldsymbol{u}) = -\frac{1}{\sqrt{n}} \sum_{t=1}^n \boldsymbol{X}_t^T \boldsymbol{u}[I(\varepsilon_t > 0) - I(\varepsilon_t < 0)] + 2\tau(\boldsymbol{u}). \qquad (8)$$

It is easy to see that we can approximate $a_n(\widehat{\boldsymbol{\phi}}_n - \boldsymbol{\phi})$ by the minimizer of Z_n^*. For example, if $\psi(t) = \lambda t$ (for some $\lambda > 0$) and $a_n = \sqrt{n}$, it follows that $\tau(\boldsymbol{u}) = \lambda \boldsymbol{u}^T C \boldsymbol{u}/2$ and so we can approximate $\sqrt{n}(\widehat{\boldsymbol{\phi}}_n - \boldsymbol{\phi})$ by

$$\frac{1}{2\lambda\sqrt{n}} \sum_{t=1}^{n} C^{-1} \boldsymbol{X}_t [I(\varepsilon_t > 0) - I(\varepsilon_t < 0)].$$

More generally, we have

$$a_n(\widehat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}) \approx \boldsymbol{h}^{-1}(\boldsymbol{W}_n/2)$$

where $\boldsymbol{h}(\boldsymbol{u})$ is the gradient of $\tau(\boldsymbol{u})$, $\boldsymbol{h}^{-1}$ its inverse and

$$\boldsymbol{W}_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{X}_t [I(\varepsilon_t > 0) - I(\varepsilon_t < 0)]. \tag{9}$$

(Typically, $\boldsymbol{h}(\boldsymbol{u}) = E[\boldsymbol{X}_t \psi(\boldsymbol{u}^T \boldsymbol{X}_t)]$.) Theorems which deal with the asymptotic behaviour of this approximation error are commonly known as Bahadur-Kiefer theorems due to their connection with the work of Bahadur (1966) and Kiefer (1967) for sample quantiles. What will be proved below is an "in distribution" (as opposed to "almost sure") Bahadur-Kiefer theorem. The following lemma will be useful in determining the asymptotic behaviour of the approximation error.

Lemma 3 *Define*

$$g_n(\boldsymbol{u}) = -\boldsymbol{x}_n^T \boldsymbol{u} + \rho_n(\boldsymbol{u}) \qquad h_n(\boldsymbol{u}) = -\boldsymbol{x}_n^T \boldsymbol{u} + \rho(\boldsymbol{u})$$

and let $\boldsymbol{u}_n = \mathrm{argmin}(g_n)$, $\boldsymbol{v}_n = \mathrm{argmin}(h_n)$. *Suppose that*

(i) $\boldsymbol{x}_n \to \boldsymbol{x}_0$;

(ii) $\boldsymbol{u}_n - \boldsymbol{v}_n \to 0$;

(iii) for any t, $\boldsymbol{u}$ *and* $\boldsymbol{w}$,

$$\rho_n(\boldsymbol{u} + t\boldsymbol{w}) - \rho_n(\boldsymbol{u}) = \int_0^t \boldsymbol{w}^T \boldsymbol{\psi}_n(\boldsymbol{u} + s\boldsymbol{w})\, ds$$

and

$$\rho(\boldsymbol{u} + t\boldsymbol{w}) - \rho(\boldsymbol{u}) = \int_0^t \boldsymbol{w}^T \boldsymbol{\psi}(\boldsymbol{u} + s\boldsymbol{w})\, ds$$

for some functions $\{\boldsymbol{\psi}_n\}$ *and* $\boldsymbol{\psi}$ *where* $\boldsymbol{\psi}$ *is one-to-one.*

(iv) $\boldsymbol{v}_0 = \boldsymbol{\psi}^{-1}(\boldsymbol{x}_0)$ *exists and for some* $\alpha > 0$

$$\|\boldsymbol{\psi}(\boldsymbol{u}) - \boldsymbol{\psi}(\boldsymbol{v})\| \le k\|\boldsymbol{u} - \boldsymbol{v}\|^\alpha$$

for all $\boldsymbol{u}$, $\boldsymbol{v}$ *in a neighbourhood of* $\boldsymbol{v}_0$;
(v) For some sequence $\{b_n\}$ *with* $b_n \to \infty$ *and any compact set* K,

$$\sup_{u \in K} \|b_n(\boldsymbol{\psi}_n(\boldsymbol{u}) - \boldsymbol{\psi}(\boldsymbol{u})) - \boldsymbol{d}_0(\boldsymbol{u})\| \to \boldsymbol{0}$$

where $\boldsymbol{d}_0$ *is a continuous function.*
Then

$$b_n(\boldsymbol{\psi}(\boldsymbol{u}_n) - \boldsymbol{\psi}(\boldsymbol{v}_n)) \to -\boldsymbol{d}_0(\boldsymbol{v}_0)$$

where $\boldsymbol{v}_0 = \boldsymbol{\psi}^{-1}(\boldsymbol{x}_0)$.

A proof of Lemma 3 will not be given here. Note that if the function $\boldsymbol{\psi}(\boldsymbol{u})$ has continuous partial derivatives at $\boldsymbol{u} = \boldsymbol{v}_0$ then under the conditions of Lemma 3 we have

$$b_n(\boldsymbol{u}_n - \boldsymbol{v}_n) \to -H_\rho^{-1}(\boldsymbol{v}_0)\boldsymbol{d}_0(\boldsymbol{v}_0)$$

provided $H_\rho(\boldsymbol{u})$, the Hessian of ρ, is invertible at $\boldsymbol{u} = \boldsymbol{v}_0$.

Lemma 3 will be applied to sequences of random elements by appealing to a Skorokhod-type arguments (see, for example, van der Vaart and Wellner, 1996) to construct almost surely convergent sequences. To do this, we will define a space $B_r(R^d)$ of locally bounded R^r-valued functions defined on R^d. (By "locally bounded" we mean bounded on compact sets.) If $\{\boldsymbol{g}_n\}$ and $\boldsymbol{g}$ are elements of $B_r(R^d)$ then we will say that $\{\boldsymbol{g}_n\}$ converges to $\boldsymbol{g}$ if

$$\sup_{u \in K} \|\boldsymbol{g}_n(\boldsymbol{u}) - \boldsymbol{g}(\boldsymbol{u})\| \to 0$$

for all compact subsets K of R^d. A possible metric for this topology is

$$d(\boldsymbol{g}, \boldsymbol{h}) = \sum_{k=1}^{\infty} \min(1, d_k(\boldsymbol{g}, \boldsymbol{h}))2^{-k}$$

where

$$d_k(\boldsymbol{g}, \boldsymbol{h}) = \sup_{\|u\| \le k} \|\boldsymbol{g}(\boldsymbol{u}) - \boldsymbol{h}(\boldsymbol{u})\|.$$

We also define $C_r(R^d)$ to be the space of R^r-valued continuous functions on R^d; $C_r(R^d)$ is a separable subset of $B_r(R^d)$. If $\{\boldsymbol{D}_n\}$ and $\boldsymbol{D}$ are random elements of $B_r(R^d)$ such that $\boldsymbol{D}_n \to_d \boldsymbol{D}$ and $\boldsymbol{D}$ is (with probability 1) a random element of $C_r(R^d)$ then it is possible find almost surely convergent representations of $\{\boldsymbol{D}_n\}$ and $\boldsymbol{D}$.

(A4) For each compact set K,

$$\sup_{u\in K} n^{1/4}\left\|E[\boldsymbol{X}_t r_n(\boldsymbol{u}^T\boldsymbol{X}_t)]\right\| \to 0$$

as $n \to \infty$.

(A5) For each compact set K, we have

$$\sup_{u\in K}\left|\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{X}_t^T\boldsymbol{X}_t r_n(\boldsymbol{u}^T\boldsymbol{X}_t)\right| \to_p 0$$

as $n \to \infty$.

(A6) For each $\boldsymbol{u}$, $E\left[\boldsymbol{X}_t^T\boldsymbol{X}_t|\psi(\boldsymbol{u}^T\boldsymbol{X}_t)|\right]$ is finite.

Theorem 4 *Suppose that $\{Y_t\}$ is an AR(p) process satisfying (1) with Z_n and Z_n^* defined as in (7) and (8). Then under conditions (A1)-(A6), we have*

$$n^{1/4}(Z_n(\boldsymbol{u}) - Z_n^*(\boldsymbol{u})) \to_d V(\boldsymbol{u}) \quad \text{as } n \to \infty$$

on $C_1(R^{p+1})$ where

$$V(\boldsymbol{u}+t\boldsymbol{w}) - V(\boldsymbol{u}) = 2\int_0^t \boldsymbol{w}^T\boldsymbol{D}(\boldsymbol{u}+s\boldsymbol{w})\,ds$$

and $\boldsymbol{D}(\boldsymbol{u})$ is a zero mean Gaussian process with $\boldsymbol{D}(\boldsymbol{0}) = \boldsymbol{0}$ and

$$E\left[(\boldsymbol{D}(\boldsymbol{u}) - \boldsymbol{D}(\boldsymbol{v}))(\boldsymbol{D}(\boldsymbol{u}) - \boldsymbol{D}(\boldsymbol{v}))^T\right] = E\left[\boldsymbol{X}_t\boldsymbol{X}_t^T|\psi(\boldsymbol{u}^T\boldsymbol{X}_t) - \psi(\boldsymbol{v}^T\boldsymbol{X}_t)|\right].$$

Proof: Define

$$V_n(\boldsymbol{u}) = n^{1/4}(Z_n(\boldsymbol{u}) - Z_n^*(\boldsymbol{u}))$$

and note that $V_n(\boldsymbol{0}) = 0$ for all n. We also have

$$V_n(\boldsymbol{u}+t\boldsymbol{w}) - V_n(\boldsymbol{u}) = 2\int_0^t \boldsymbol{w}^T\boldsymbol{D}_n(\boldsymbol{u}+s\boldsymbol{w})\,ds$$

where

$$\begin{aligned}\boldsymbol{D}_n(\boldsymbol{u}) = \\ &\tfrac{1}{n^{1/2}}\textstyle\sum_{t=1}^{n}[n^{1/4}\boldsymbol{X}_t(I(\varepsilon_t \le \boldsymbol{u}^T\boldsymbol{X}_t/a_n) \\ &-I(\varepsilon_t \le 0)) - n^{-1/4}E[\boldsymbol{X}_t\psi(\boldsymbol{u}^T\boldsymbol{X}_t)]]\end{aligned}$$

since our assumptions imply that the gradient of $\tau(\boldsymbol{u}) = E[\Psi(\boldsymbol{u}^T\boldsymbol{X}_t)]$ is $E[\boldsymbol{X}_t\psi(\boldsymbol{u}^T\boldsymbol{X}_t)]$. Clearly, $\boldsymbol{D}_n(\boldsymbol{0}) = \boldsymbol{0}$ for all n and applying an appropriate martingale central limit theorem (Hall and Heyde, 1980), it follows that the finite dimensional distributions of $\boldsymbol{D}_n$ converge to those of $\boldsymbol{D}$. It is also straightforward to verify that on each compact set K,

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty} P\left(\sup_{\|u-v\|<\delta; u,v\in K} \|\boldsymbol{D}_n(\boldsymbol{u}) - \boldsymbol{D}_n(\boldsymbol{v})\| > \epsilon\right) = 0$$

for every $\epsilon > 0$ by using an appropriate moment condition. Hence $\boldsymbol{D}_n \to_d \boldsymbol{D}$ on $B_{p+1}(R^{p+1})$ and the conclusion follows. □

Theorem 5 *Assume the conditions of Theorem 4 and let $\boldsymbol{h}(\boldsymbol{u})$ be the gradient of $\tau(\boldsymbol{u})$ with inverse $\boldsymbol{h}^{-1}$. If $\boldsymbol{U}$ minimizes Z and*

$$\|\boldsymbol{h}(\boldsymbol{u}) - \boldsymbol{h}(\boldsymbol{v})\| \leq k\|\boldsymbol{u} - \boldsymbol{v}\|^{\alpha} \quad (\alpha > 0)$$

for all $\boldsymbol{u}$, $\boldsymbol{v}$ in a neighbourhood of $\boldsymbol{U}$ (k and α may depend on $\boldsymbol{U}$) then

$$n^{1/4}\left(\boldsymbol{h}(a_n(\widehat{\boldsymbol{\phi}}_n - \boldsymbol{\phi})) - \frac{\boldsymbol{W}_n}{2}\right) \to_d -\boldsymbol{D}(\boldsymbol{h}^{-1}(\boldsymbol{W}/2))$$

as $n \to \infty$ where $\boldsymbol{W}_n$ is defined in (9), $\boldsymbol{D}$ is the Gaussian process defined in Theorem 4 and $\boldsymbol{W}$ is a $(p+1)$-variate Normal random vector (independent of $\boldsymbol{D}$) with mean $\boldsymbol{0}$ and covariance matrix C.

Proof: Let $\boldsymbol{U}_n = a_n(\widehat{\boldsymbol{\phi}}_n - \boldsymbol{\phi})$ minimize Z_n. Then it is easy to verify that

$$\left(\boldsymbol{U}_n, n^{1/4}(Z_n - Z_n^*)\right) \to_d \left(\boldsymbol{h}^{-1}(W/2), V\right)$$

as $n \to \infty$ on the space $R^{p+1} \times B_1(R^{p+1})$ where W and V are independent. Since the limit is concentrated on a separable subset of $R^{p+1} \times B_1(R^{p+1})$ (namely $R^{p+1} \times C_1(R^{p+1})$), we can construct a probability space and almost surely convergent versions of $\{\boldsymbol{U}_n\}$ and $\{n^{1/4}(Z_n - Z_n^*)\}$. The conclusion follows by applying Lemma 3 to each convergent sequence. □

If $\boldsymbol{h}(\boldsymbol{u})$ is one-to-one (with inverse $\boldsymbol{h}^{-1}$) and continuously differentiable then it follows that (under the conditions of Theorem 5)

$$n^{1/4}\left[a_n(\widehat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}) - \boldsymbol{h}^{-1}(\boldsymbol{W}_n/2)\right] \to_d -H^{-1}(\boldsymbol{h}^{-1}(\boldsymbol{W}/2))\boldsymbol{D}(\boldsymbol{h}^{-1}(\boldsymbol{W}/2))$$

provided that $H(\boldsymbol{u})$, the Hessian of τ, is invertible outside of a set of Lebesgue measure 0 in R^{p+1}. This suggests the asymptotic expansion

$$\begin{aligned} a_n(\widehat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}) &= \boldsymbol{h}^{-1}(\boldsymbol{W}_n/2) \\ &\quad - \frac{1}{n^{1/4}} H^{-1}(\boldsymbol{h}^{-1}(\boldsymbol{W}_n/2))\boldsymbol{D}(\boldsymbol{h}^{-1}(\boldsymbol{W}_n/2)) + o_p(n^{-1/4}). \end{aligned}$$

(As before, $\boldsymbol{W}_n$ is defined as in (9).) Whether this expansion is particularly useful is an open question.

Evaluating the limiting distribution in Theorem 5 is tedious but not overly difficult (provided, of course, that everything about $\{Y_t\}$ is known). For a given $\boldsymbol{u}$, $\boldsymbol{D}(\boldsymbol{u})$ is $(p+1)$-variate Normal with mean $\mathbf{0}$ and covariance matrix

$$K(\boldsymbol{u}) = E[\boldsymbol{X}_t\boldsymbol{X}_t^T|\psi(\boldsymbol{u}^T\boldsymbol{X}_t)|].$$

If $K(\boldsymbol{u})$ is positive definition for $\boldsymbol{u}$ outside of a set of Lebesgue measure 0 then since $\boldsymbol{W}$ is independent of $\boldsymbol{D}$, it follows that the density of $-\boldsymbol{D}(\boldsymbol{h}^{-1}(\boldsymbol{W}/2))$ is

$$f_1(\boldsymbol{x}) = \frac{1}{|C|^{1/2}\pi^{p+1}}\int \frac{|H(\boldsymbol{u})|}{|K(\boldsymbol{u})|^{1/2}}\exp\left[-\frac{1}{2}\gamma_1(\boldsymbol{x},\boldsymbol{u})\right] d\boldsymbol{u}$$

where

$$\gamma_1(\boldsymbol{x},\boldsymbol{u}) = \boldsymbol{x}^T K^{-1}(\boldsymbol{u})\boldsymbol{x} + 4\boldsymbol{h}(\boldsymbol{u})^T C^{-1}\boldsymbol{h}(\boldsymbol{u})$$

and the integration is over R^{p+1} with $|\cdot|$ denoting determinant. Likewise the density of $-H^{-1}(\boldsymbol{h}^{-1}(\boldsymbol{W}/2))\boldsymbol{D}(\boldsymbol{h}^{-1}(\boldsymbol{W}/2))$ is

$$f_2(\boldsymbol{x}) = \frac{1}{|C|^{1/2}\pi^{p+1}}\int \frac{|H(\boldsymbol{u})|^2}{|K(\boldsymbol{u})|^{1/2}}\exp\left[-\frac{1}{2}\gamma_2(\boldsymbol{x},\boldsymbol{u})\right] d\boldsymbol{u}$$

where

$$\gamma_2(\boldsymbol{x},\boldsymbol{u}) = \boldsymbol{x}^T H(\boldsymbol{u})K^{-1}(\boldsymbol{u})H(\boldsymbol{u})\boldsymbol{x} + 4\boldsymbol{h}(\boldsymbol{u})^T C^{-1}\boldsymbol{h}(\boldsymbol{u}).$$

In the following example, we derive the density f_2 in a simple case.

Example 2. Let $Y_t = \varepsilon_t$ where $\{\varepsilon_t\}$ are i.i.d. random variables with distribution function F satisfying $F'(0) = \lambda > 0$. Suppose that we estimate only the parameter ϕ_1 of an AR(1) model; call this estimator $\widehat{\phi}_n$. We then have

$$\tau(u) = \frac{1}{2}\lambda\sigma^2u^2$$

where $\sigma^2 = E(\varepsilon_t^2)$. We also have $C = \sigma^2$ and $H(u) = \tau''(u) = \lambda\sigma^2$. Finally, $K(u) = \lambda\gamma|u|$ where $\gamma = E[|\varepsilon_t|^3]$. It follows from Theorems 4 and 5 that

$$n^{1/4}\left(\sqrt{n}\widehat{\phi}_n - \frac{1}{2\lambda\sqrt{n}}\sum_{t=1}^{n} Y_{t-1}[I(\varepsilon_t > 0) - I(\varepsilon_t < 0)]\right) \to_d S$$

where S has density

$$f_S(x) = \frac{\lambda^{3/2}\sigma^2}{\pi\gamma^{1/2}}\int_{-\infty}^{\infty} |u|^{-1/2}\exp\left[-\frac{1}{2}\left(\frac{\lambda\sigma^4x^2}{\gamma|u|} + 4\lambda^2\sigma^2u^2\right)\right] du.$$

S has a symmetric distribution with moment generating function

$$E[\exp(tS)] = 2\exp\left(\frac{\gamma^2 t^4}{32\lambda^4\sigma^{10}}\right)\Phi\left(\frac{\gamma t^2}{2\lambda^2\sigma^5}\right)$$

where Φ is the standard Normal distribution function.

4 Final comments

In this paper, we have derived first- and second-order limiting distributions for the L_1-estimators of the parameters of an AR(p) process under fairly general conditions on the error distribution. From a statistical point of view, the fact that the asymptotic behaviour of the L_1-estimators is so sensitive to the behaviour of $F(x)$ for x close to 0 is somewhat troubling. One possible non-parametric approach to estimating the sampling distribution of $\widehat{\boldsymbol{\phi}}_n$ is to bootstrap the AR(p) process by sampling with replacement from the residuals $e_t = Y_t - \boldsymbol{X}_t^T\widehat{\boldsymbol{\phi}}_n$. However, it is possible to show that, asymptotically, this bootstrap procedure is correct to first order only if $\psi(t)$ is a linear function and is never correct to second order. This is similar to the results of Hall and Martin (1988) and Huang *et al.* (1996) for sample quantiles of i.i.d. random variables. However, other approaches to bootstrapping time series, such as frequency domain bootstrapping, may prove to be more fruitful in this problem.

It may also be possible to exploit Lemma 3 to obtain an "almost sure" Bahadur-Kiefer representation. Arcones (1996a, 1996b) and He and Shao (1996) derive such representations for L_p-estimators in linear regression models. However, these papers assume that $F(x)$, the distribution function of the errors, is linear in a neighbourhood of $x = 0$. Using the notation of Theorems 4 and 5, we can conjecture that, under appropriate regularity conditions,

$$(n/\ln(\ln(n)))^{1/4}\left(\boldsymbol{h}(b_n(\widehat{\boldsymbol{\phi}}_n - \boldsymbol{\phi})) - \frac{\boldsymbol{W}_n}{2\sqrt{\ln(\ln(n))}}\right) = O(1) \qquad (10)$$

with probability 1 where b_n satisfies the condition

$$\lim_{n\to\infty}\sqrt{n/\ln(\ln(n))}(F(t/b_n) - F(0)) = \psi(t)$$

and the set of limit points of the left hand side of (10) is non-trivial.

Acknowledgements

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. The author would like to thank a referee for some useful comments.

References

[1] Arcones, M.A. (1996a). The Bahadur-Kiefer representation of L_p regression estimators. *Econometric Theory* **12**, 257-283.

[2] Arcones, M.A. (1996b). Second order representations of the least absolute deviation regression estimator. Unpublished manuscript.

[3] Bahadur, R.R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37**, 577-580.

[4] de Haan, L. and E. Taconis-Haantjes (1979). On Bahadur's representation of sample quantiles. *Ann. Inst. Statist. Math.* **31**, 299-308.

[5] Geyer, C.J. (1996). On the asymptotics of convex stochastic optimization. Unpublished manuscript.

[6] Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and its Application.* New York: Academic Press.

[7] Hall, P. and M.A. Martin (1988). Exact convergence rate of bootstrap quantile variance estimator. *Probability theory and related fields* **80**, 261-268.

[8] He, X. and Q.-M. Shao (1996). A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. *Ann. Statist.* **24**, 2608-2631.

[9] Huang, J.S., P.K. Sen and J. Shao (1996). Bootstrapping a sample quantile when the density has a jump. *Statistica Sinica* **6**, 299-309.

[10] Kiefer, J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Statist.* **38**, 1323-1342.

[11] Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7**, 186-199.

[12] Smirnov, N.V. (1952). Limit distributions for the terms of a variational series. *American Mathematical Society Translation Series (1)* **11**, 82-143.

[13] van der Vaart, A.W. and J.A. Wellner (1996). *Weak Convergence and Empirical Processes with Applications to Statistics.* New York: Springer.

[14] Wang, L. and J. Wang (1996). Limit distribution of L_1-estimators for time series: stationary linear models. Unpublished manuscript.