Ann. Inst. Statist. Math.
Vol. 56, No. 2, 251–264 (2004)

ASYMPTOTICS OF ESTIMATES IN CONSTRAINED NONLINEAR REGRESSION WITH LONG-RANGE DEPENDENT INNOVATIONS

LIHONG WANG

Department of Mathematics, Nanjing University, Nanjing 210093, China

(Received April 30, 2002; revised March 10, 2003)

Abstract. The purpose of this paper is to investigate the asymptotic properties of the least squares estimates (L_2-estimates) and the least absolute deviation estimates (L_1-estimates) of the parameters of a nonlinear regression model subject to a set of equality and inequality restrictions, which has a long-range dependent stationary process as its stochastic errors. Then we will compare the asymptotic relative efficiencies of the above estimators.

Key words and phrases: Asymptotic efficient, asymptotic property, least absolute deviation estimation, least squares estimation, long-range dependence, equality and inequality constraints, nonlinear regression.

1. Introduction

Recently, there is a growing need for the statistical methodologies based on long-range dependence data. Long range dependence appears naturally in various areas, especially hydrology, economics, geophysics, communications, see, e.g. Yajima (1988, 1991), Beran (1992), among others. A typical behavior of long-range dependent sequence is that the covariance of this sequence decreases to zero like a power of lag as the lag tends to infinity, but their absolute sum diverges.

Many authors such as Yajima (1988, 1991), Robinson (1995), Dahlhaus (1995), Robinson and Hidalgo (1997) studied the asymptotic properties of (weighted or generalized) L_2-estimation for certain forms of linear regression models in the presence of long range dependence in errors. Koul and Mukherjee (1993), Giraitis *et al.* (1996) also considered the asymptotic properties of various robust estimates in linear regressions involving long-range dependence. For unconstrained nonlinear regression with long-range Gaussian subordinated errors, the asymptotic behavior of some robust estimates, was investigated in Koul (1996).

Recently, constrained regression problems have been dealt with by many authors. One can see Liew (1976), Nagaraj and Fuller (1991), Wang (1996) for the motivations of making restrictions on the parameters and for the various approaches to deriving the asymptotic distributions of L_2-estimators for constrained regressions with independent innovations. Typically the restrictions on the choice of the parameters could be much complex ones including certain mathematical relations between the parameters that need to be estimated. Classical techniques, which can still be used to handle least-squares estimation with linear equality constraints on the parameters for example, break down if there are inequality constraints or a non-differentiable criterion function. Usually, the relationship between the samples and the estimates must be found by solving an optimization problem. We may use mathematical programming techniques such as epigraph convergence in optimization to get constrained estimates as explored by Wang (1996).

Koul and Mukherjee (1993) and Koul (1996) showed that some robust estimates including L_1-estimates are asymptotically equivalent to the L_2-estimates in the case of Gaussian long-range dependent errors for unconstrained models. Then one may ask whether this conclusion hold or not in constrained case. Therefore it is of interest and importance to examine the asymptotic behaviors of L_1-estimators in constrained nonlinear regression models when errors are functions of long-range dependent Gaussian random variables and compare its asymptotic properties with those of L_2-estimators.

Specifically, in this paper we shall consider the following nonlinear regression model with errors whose unknown distribution function (d.f.) is F.

$$y_i = f(x_i, \theta) + e_i, \qquad i \geq 1, \tag{1.1}$$

where $\theta \in R^d$ is the unknown parameter to be estimated, $e_i = G(\eta_i)$, $i \geq 1$, G can be viewed as $F^{-1}\Phi$ (Φ is the d.f. of normal $N(0,1)$ and $F^{-1}(u) = \inf\{x : F(x) \geq u\}$, $0 \leq u \leq 1$). We will assume that $\eta_1, \eta_2, \ldots$ is a sequence of stationary long-range dependent Gaussian random variables.

Moreover, in many practical cases, we may have some prior knowledge about the parameters. For example, this occurs when the experimenter has a strong belief that the parameters lie in some irregular region of R^d. For generality, let the prior information on θ be given by

$$\bar{S} = \{\theta : g_i(\theta) \leq 0, i = 1, \ldots, m; h_j(\theta) = 0, j = 1, \ldots, p\}. \tag{1.2}$$

Here $g_i(\theta)$, $h_j(\theta)$ are functional restrictions (may be nonlinear) on θ.

For instance, the constraints can be of the form

$$A\theta \leq C, \qquad B\theta = D,$$

where $A_{m\times d}$, $B_{p\times d}$, $C_{m\times 1}$, $D_{p\times 1}$ are given matrices. See Shapiro (1988) and the following example.

Example 1.1. (Heywood cases in factor analysis, Joreskog (1969)) The model for confirmative factor analysis can be simplified as

$$y = f(x, \Psi) + e,$$

where $f(x, \Psi) = \log|\Psi| - \log|x|$, Ψ is a $m \times m$ matrix and its unknown diagonal elements Ψ_{ii} should be nonnegative. So we have the restrictions

$$\Psi_{ii} \geq 0, \qquad i = 1, \ldots, m.$$

Example 1.2. (Nonlinear constraints case) In connection with the maximum likelihood estimation (MLE), the case of parameter restrictions in the form of smooth nonlinear functions was studied by Aitchison and Silvey (1958). Moreover, in the pioneering paper by Huber (1967), nonstandard sufficient conditions were given under which the MLE is an asymptotically optimal estimator. For certain practical implementation, if we remove these assumptions to get result valid, $\bar{S}$, a proper subset of R^d, can be identified from these nonstandard sufficient conditions and extraneous considerations and then actually a restricted MLE are derived for a nonlinear-constrained regression like model (1.1). As an example, $\bar{S}$ can be

$$\bar{S} = \{\theta : \theta_1^2 + \theta_2^2 + \cdots + \theta_d^2 - 1 \leq 0; \theta_2 \leq \theta_3 \leq \cdots \leq \theta_d; h(\theta_1 - \theta_2) = 0\}.$$

For more details, see Silvey (1959) and Sen (1979).

Given observations (x_i, y_i), $i = 1, \ldots, n$, the L_1-estimator, denoted by $\hat{\theta}^{(1)}$, is the optimal solution of the following minimizing problem:

$$\min_{\theta \in \bar{S}} \sum_{i=1}^{n} |y_i - f(x_i, \theta)|, \tag{1.3}$$

where $\bar{S}$ is defined in (1.2).

The L_2-estimator, denoted by $\hat{\theta}^{(2)}$, is the optimal solution of

$$\min_{\theta \in \bar{S}} \sum_{i=1}^{n} |y_i - f(x_i, \theta)|^2. \tag{1.4}$$

We will give conditions on the design in order that $a_{n1}(\hat{\theta}^{(1)} - \theta_0)$, $a_{n2}(\hat{\theta}^{(2)} - \theta_0)$ converge in distribution for some suitable sequences a_{n1}, a_{n2}, where θ_0 is the true value of θ. To do this, we use equivalent forms of problems (1.3) and (1.4):

$$\begin{aligned} &\min \ \frac{a_{n1}^2}{n} \sum_{i=1}^{n} \{|e_i - T_i(v)| - |e_i|\} \\ &\text{s.t.} \ \ v \in S_{n1} \end{aligned} \tag{1.5}$$

and

$$\begin{aligned} &\min \ \frac{a_{n2}^2}{n} \sum_{i=1}^{n} \{|e_i - T_i(w)|^2 - |e_i|^2\} \\ &\text{s.t.} \ \ w \in S_{n2} \end{aligned} \tag{1.6}$$

where $S_{nr} = \{v : g_i(\theta_0 + a_{nr}^{-1} v) \leq 0, i = 1, \ldots, m; h_j(\theta_0 + a_{nr}^{-1} v) = 0, j = 1, \ldots, p\}$, $r = 1, 2$.

Denote the objective functions of (1.5) and (1.6) by $Q_n(v)$ and $V_n(w)$, the optimal solutions by $\hat{v}$, $\hat{w}$ respectively, here $v = a_{n1}(\theta - \theta_0)$, $T_i(v) = f(x_i, \theta_0 + a_{n1}^{-1} v) - f(x_i, \theta_0)$; $w = a_{n2}(\theta - \theta_0)$, $T_i(w) = f(x_i, \theta_0 + a_{n2}^{-1} w) - f(x_i, \theta_0)$. Obviously, $\hat{v}$, $\hat{w}$ are simply $a_{n1}(\hat{\theta}^{(1)} - \theta_0)$ and $a_{n2}(\hat{\theta}^{(2)} - \theta_0)$.

Our aim is to establish the asymptotic behaviors of $\hat{v}$ and $\hat{w}$. We briefly describe the main idea. First we show that, under some regularity conditions, $Q_n(v)$, $V_n(w)$ converge uniformly in distribution to some functions $Q(v)$, $V(w)$ respectively, then we prove that $a_{n1}(\hat{\theta}^{(1)} - \theta_0)$ converges in distribution to the optimal solution of the following program:

$$\begin{aligned} &\min \ Q(v) \\ &\text{s.t.} \ \ v \in S \end{aligned} \tag{1.7}$$

where

$$S = \{v : (\nabla g_i(\theta_0))'v \le 0, i \in J; (\nabla h_j(\theta_0))'v = 0, j = 1, \dots, p\}$$

with $J = \{i : g_i(\theta_0) = 0, i = 1, \dots, m\}$ and $a_{n2}(\hat{\theta}^{(2)} - \theta_0)$ converges in distribution to the optimal solution of the program:

$$\min_{w \in S} V(w). \tag{1.8}$$

Here $\nabla g_i(\theta_0)$, $\nabla h_j(\theta_0)$ are the gradient vectors of $g_i(\theta)$, $h_j(\theta)$ with respect to θ at $\theta = \theta_0$. This technique is also used in Wang (1996).

We shall see in Section 2 that the limiting law of $a_{n1}(\hat{\theta}^{(1)} - \theta_0)$, which is formed by a multiple Itô-Wiener integral, is usually non normal, if the covariance of η_i decays as a regularly varying function. The result is stated in Theorem 2.2. Section 3 is devoted to the asymptotic distribution of $a_{n2}(\hat{\theta}^{(2)} - \theta_0)$, which is presented in Theorem 3.2. Furthermore, we discuss the asymptotic efficiencies of $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$ in Section 4. We shall prove that L_1-estimate is asymptotically equivalent to L_2-estimate if the errors are Gaussian. This result is similar to what happens in the unconstrained case. But it is in complete contrast with the i.i.d. errors case. Finally, we show that L_1-estimate is much more efficient than L_2-estimate at the double exponential error distribution, the logistic error distribution.

In the sequel, C stands for a constant whose value is not of interest and may vary from line to line.

2. Limiting distribution of L_1-estimate $\hat{\theta}^{(1)}$

To proceed further, let H_k, $k = 1, 2, \dots$ be the Hermite polynomials, where

$$H_k(x) = (-1)^k e^{x^2/2} \frac{d^k}{dx^k} e^{-x^2/2}.$$

Set

$$u_k = E\{\text{sign}(G(\eta_1))H_k(\eta_1)\}, \qquad k \ge 1. \tag{2.1}$$

Then $\text{sign}(e_i)$ has an Hermite expansion given by

$$\text{sign}(e_i) = \text{sign}(G(\eta_i)) = \sum_{k=\tau_1}^{\infty} \frac{u_k}{k!} H_k(\eta_i).$$

The Hermite rank of $\text{sign}(e_i)$ is

$$\tau_1 = \inf\{k \ge 1 : u_k \ne 0\}. \tag{2.2}$$

For more details on Hermite expansion, see Taqqu (1975, 1979).

We assume the following assumptions:

(C1) $\eta_1, \eta_2, \dots$ are stationary, mean zero, unit variance Gaussian random variables. $e_i = G(\eta_i)$ has finite variance, and the d.f. F of e_i is symmetric around median zero and F has a density function F' such that F' is positive and continuous at 0;

(C2) $\gamma(k) = E(\eta_1\eta_{k+1}) = k^{-\alpha}L(k)$, $k \geq 0$, where $L(\cdot)$ is a slowly varying function (at infinity) in the sense that

$$\lim_{s\to\infty} \frac{L(st)}{L(s)} = 1 \qquad \text{for every} \quad t \in (0, \infty)$$

and $0 < \alpha < 1/\tau_1$, τ_1 is defined in (2.2);

(C3) $f(x_i, \theta)$, $i = 1, \dots, n$ are continuously differentiable in θ at $\theta = \theta_0$ and there is a neighborhood U of θ_0 such that for all θ in U it holds that

$$f(x_i, \theta) - f(x_i, \theta_0) - (\nabla f(x_i, \theta_0))'(\theta - \theta_0) = r_i(\theta)\|\theta - \theta_0\|^2,$$

where $\nabla f(x_i, \theta_0) = (\nabla f^{(1)}(x_i, \theta_0), \dots, \nabla f^{(d)}(x_i, \theta_0))'$ is the gradient vector of $f(x_i, \theta)$ at $\theta = \theta_0$. $\|\cdot\|$ denotes the Euclidean norm; and $\max_i |r_i(\theta)| < \infty$ uniformly on U;

(C4) The sequence of matrices

$$n^{-1} \sum_{i=1}^{n} \nabla f(x_i, \theta_0)(\nabla f(x_i, \theta_0))'$$

has a limit matrix $\boldsymbol{K}$, which is positive definite and $\max_i \|\nabla f(x_i, \theta_0)\| < \infty$;

(C5) $g_i(\theta)$, $i = 1, \dots, m$; $h_j(\theta)$, $j = 1, \dots, p$ are continuously differentiable in U;

(C6) $\nabla g_i(\theta_0)$, $i \in J$; $\nabla h_j(\theta_0)$, $j = 1, \dots, p$ are linearly independent.

Before presenting the results of this section, let us briefly discuss the assumptions. Assumption (C1) is imposed to establish the asymptotics of L_1-estimator, while Assumption (C2) means that e_i is a long-range dependent error sequence. The purpose of imposing Assumptions (C3)–(C6) is to study the asymptotics of estimators in nonlinear constrained regression problems. Therefore we may say that Assumptions (C1)–(C6) are quite natural and mild. Similar assumptions as (C1)–(C4) were also imposed in Koul (1996).

Now let

$$a_{n1}^2 = n^{\tau_1\alpha} L^{-\tau_1}(n).$$

We have the following

THEOREM 2.1. *Under Assumptions* (C1)–(C4), *it follows that*

$$Q_n(v) \to Q(v) = F'(0)v'\boldsymbol{K}v - v'\zeta$$

in distribution uniformly for all v, where ζ is a d-dimensional real-valued, mean zero random vector,

$$\zeta = \frac{u_{\tau_1}}{\tau_1!} D^{-\tau_1/2} \int_{R^{\tau_1}} K_0(x_1, \dots, x_{\tau_1})|x_1|^{(\alpha-1)/2} \cdots |x_{\tau_1}|^{(\alpha-1)/2} dW(x_1) \cdots dW(x_{\tau_1}),$$

where $D = 2\Gamma(\alpha)\cos(\alpha\pi/2)$, u_{τ_1} is defined in (2.1) *and $K_0(x_1, \dots, x_{\tau_1})$ is a d-dimensional vector with*

$$\begin{aligned} K_{0j}(x_1, \dots, x_{\tau_1}) &= \lim_{n\to\infty} K_{nj}(x_1, \dots, x_{\tau_1}) \\ &:= \lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} \nabla f^{(j)}(x_t, \theta_0) e^{i(x_1+\cdots+x_{\tau_1})/nt}, \qquad j = 1, \dots, d. \end{aligned}$$

W is the random spectrum of the Gaussian white-noise process, and the covariance matrix of ζ *is* $\frac{u_{\tau_1}^2}{\tau_1!}\Sigma(\tau_1,\alpha)$, *which is positive definite with*

$$\Sigma(\tau_1,\alpha)=\lim_{n\to\infty} n^{\tau_1\alpha-1}L^{-\tau_1}(n)\sum_{i,j=1}^{n}\nabla f(x_i,\theta_0)(\nabla f(x_j,\theta_0))'(\gamma(i-j))^{\tau_1}.$$

PROOF. By Theorem 2.2 of Koul (1996) and the convexity property of the absolute value function (cf. Rockafellar (1970)), it follows easily that

$$Q_n(v)-\left\{F'(0)v'\boldsymbol{K}v-v'\frac{a_{n1}}{n}\sum_{i=1}^{n}\nabla f(x_i,\theta_0)\mathrm{sign}(e_i)\right\}\to 0$$

in probability uniformly for all v. Hence, in order to establish the uniform weak convergence of $Q_n(v)$, it suffices to show that

$$n^{\tau_1\alpha/2-1}L^{-\tau_1/2}(n)\sum_{i=1}^{n}\nabla f(x_i,\theta_0)\mathrm{sign}(e_i)\to\zeta \tag{2.3}$$

in distribution.

In fact (2.3) is the weighted version of Theorem 1′ in Dobrushin and Major (1979). Therefore the proof of Theorem 2.2 of Koul (1996), Theorem 1′ and Remark 1.1 in Dobrushin and Major (1979) implies the relation (2.3) and hence the desired result. □

THEOREM 2.2. *Suppose Assumptions* (C1)–(C6) *hold true. Then*

(i) $n^{\tau_1\alpha/2}L^{-\tau_1/2}(n)(\hat\theta^{(1)}-\theta_0)$ *is bounded in probability;*

(ii) *If the set of optimal solutions of Problem* (1.7) *is a singleton for each value of* ζ, *then* $n^{\tau_1\alpha/2}L^{-\tau_1/2}(n)(\hat\theta^{(1)}-\theta_0)$ *converges in distribution to the optimal solution* v^* *of problem* (1.7). *Furthermore,*

$$v^*=\begin{cases} M_0\zeta, & \text{if } v^*\in S_0,\\ M_{i_1,\cdots,i_k}\zeta, & \text{if } v^*\in S_{i_1,\cdots,i_k},\ k=1,2,\dots,\end{cases}$$

where

$$S_0=\{v:(\nabla g_i(\theta_0))'v<0, i\in J;(\nabla h_j(\theta_0))'v=0, j=1,\dots,p\},$$
$$S_{i_1,\cdots,i_k}=\{v:(\nabla g_i(\theta_0))'v<0, i\in J\setminus\{i_1,\dots,i_k\};(\nabla g_i(\theta_0))'v=0, i\in\{i_1,\dots,i_k\};$$
$$(\nabla h_j(\theta_0))'v=0, j=1,\dots,p\}$$

and M_0, $M_{i_1,\dots,i_k}$ *are the first blocks of the matrices* H_0^{-1}, $H_{i_1,\dots,i_k}^{-1}$ *respectively,*

$$H_0=\begin{pmatrix}2F'(0)\boldsymbol{K} & \nabla h\\ (\nabla h)' & 0\end{pmatrix},$$

$$H_{i_1,\dots,i_k}=\begin{pmatrix}2F'(0)\boldsymbol{K} & \nabla g_{i_1} & \cdots & \nabla g_{i_k} & \nabla h\\ (\nabla g_{i_1})' & 0 & \cdots & \cdots & 0\\ \vdots & \vdots & & & \vdots\\ (\nabla g_{i_k})' & 0 & \cdots & \cdots & 0\\ (\nabla h)' & 0 & \cdots & \cdots & 0\end{pmatrix},\qquad i_1,\dots,i_k\in J,$$

here ∇h stands for $(\nabla h_j(\theta_0),\ j=1,\cdots,p)$.

PROOF. Note that $v=0$ is a feasible solution of program (1.5) and $\hat{v}=a_{n1}(\hat{\theta}^{(1)}-\theta_0)$ is the optimal solution of (1.5), that is,

$$0 \geq Q_n(\hat{v}) - Q_n(0) = Q_n(\hat{v}).$$

Then to prove (i), it suffices to verify the following Claim: for any $\varepsilon>0$, there exists a constant M_ε such that with probability greater than $1-\varepsilon$, when $\|v\|>M_\varepsilon$, we have $Q_n(v)>0$.

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d > 0$ be its characteristic roots and $l_1,\ldots,l_d$ be its standardized characteristic vectors. Since ζ is a real-valued, mean zero random variable with finite positive definite covariance matrix $\mathrm{Var}(\zeta)$, it follows from Theorem 2.1 that

$$\mathrm{Var}(Q_n(v)) \to v'\mathrm{Var}(\zeta)v = \sum_{i=1}^{d} \lambda_i a_i^2 \leq \lambda_1 \|v\|^2$$

uniformly for all v, where v can be written as $v=\sum_{i=1}^{d} a_i l_i$. Hence for n large enough and any $\varepsilon>0$, by Chebyshev's inequality, we have

$$\mathrm{P}\{|Q_n(v) - E(Q_n(v))| \leq \|v\| b_\varepsilon\} \geq 1-\varepsilon$$

uniformly for all v by choosing $b_\varepsilon = \sqrt{\frac{2\lambda_1}{\varepsilon}}$. This implies that

$$\mathrm{P}\{Q_n(v) \geq -\|v\| b_\varepsilon + F'(0)v'\boldsymbol{K}v + \delta\} \geq 1-\varepsilon \tag{2.4}$$

uniformly for all v, where $\delta \to 0$ as $n \to \infty$. Again since $F'(0)\boldsymbol{K}$ is positive definite, still define $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d > 0$ and $l_1,\ldots,l_d$ be characteristic roots and standardized characteristic vectors of $F'(0)\boldsymbol{K}$, we have

$$-\|v\| b_\varepsilon + F'(0)v'\boldsymbol{K}v = -\|v\| b_\varepsilon + \sum_{i=1}^{d} \lambda_i a_i^2 > \lambda_d \|v\|^2 - \|v\| b_\varepsilon.$$

For n large enough, let M_ε, δ be chosen so that

$$\inf_{\|v\|>M_\varepsilon} \{-\|v\| b_\varepsilon + F'(0)v'\boldsymbol{K}v + \delta\} > 0.$$

This, in turn, when using (2.4), means that

$$\begin{aligned} &\mathrm{P}\left\{\inf_{\|v\|>M_\varepsilon}\{Q_n(v)\} > 0\right\} \\ &\quad \geq \mathrm{P}\left\{\inf_{\|v\|>M_\varepsilon}\{Q_n(v)\} \geq \inf_{\|v\|>M_\varepsilon}\{-\|v\| b_\varepsilon + F'(0)v'\boldsymbol{K}v + \delta\}\right\} \\ &\quad \geq 1-\varepsilon \end{aligned}$$

and hence the validity of the Claim. This completes the proof of (i).

Next we turn to verify (ii). It follows from Theorem 2 in Wang (1996) that $\lim_{n\to\infty} S_{n1} = S$ (here the convergence is in Kuratowski's sense). Then by Theorem 2.1, it is immediate to formulate a limit program of problem (1.5), that is problem (1.7):

$$\min_{v\in S} -v'\zeta + F'(0)v'\boldsymbol{K}v.$$

Applying the similar arguments as used in Section 4 of Wang (1996), we get the result that the optimal solution of problem (1.5), $\hat{v} = n^{\tau_1\alpha/2}L^{\tau_1/2}(n)(\hat{\theta}^{(1)} - \theta_0)$, converges in distribution to the optimal solution v^* of problem (1.7).

Finally we derive the distribution of v^*. For the first case that $v^* \in S_0$, by Kuhn-Tucker conditions, v^* must satisfy the equations

$$\begin{aligned} &2F'(0)\boldsymbol{K}v - \zeta + \sum_{j=1}^{p} r_j \nabla h_j(\theta_0) = 0,\\ &(\nabla h_j(\theta_0))'v = 0, \qquad j = 1,\dots,p, \end{aligned}$$

where r_j, $j = 1,\dots,p$ are the Lagrangian multipliers. The matrix form of this equation system can be written as

$$H_0\begin{pmatrix} v\\ r\end{pmatrix} = \begin{pmatrix} \zeta\\ 0\end{pmatrix}.$$

Solving this system it follows that

$$v^* = M_0\zeta. \tag{2.5}$$

Since $v^* \in S_0$ and the objective function of program (1.7) is a positive definite quadratic function, thus (2.5) is the expression of the optimal solution in S_0 of program (1.7).

If for some sample values of ζ, $M_0\zeta \notin S_0$, the optimal solution must lie on the relative boundary of S. Let us find the expression of the optimal solution on the intersection of faces $S_{i_1},\dots,S_{i_k}$. By Kuhn-Tucker conditions the optimal solution of program (1.7) located on $S_{i_1,\dots,i_k}$ must satisfy the equations

$$\begin{aligned} &2F'(0)\boldsymbol{K}v - \zeta + \sum_{t=i_1}^{i_k} \lambda_t \nabla g_t(\theta_0) + \sum_{j=1}^{p} r_j \nabla h_j(\theta_0) = 0,\\ &(\nabla g_t(\theta_0))'v = 0, \qquad t = i_1,\dots,i_k,\\ &(\nabla h_j(\theta_0))'v = 0, \qquad j = 1,\dots,p, \end{aligned} \tag{2.6}$$

where λ_t, $t = i_1,\dots,i_k$; r_j, $j = 1,\dots,p$ are the Lagrangian multipliers. The solution of system (2.6) can be expressed as

$$v^* = M_{i_1,\dots,i_k}\zeta. \tag{2.7}$$

As program (1.7) is a convex program, the optimal solution of program (1.7) on $S_{i_1,\dots,i_k}$ has the expression (2.7). Thus the proof of Theorem 2.2 is completed. □

Remark 2.1. For the case of $\tau_1 = 1$, from the Hermite expansion of $\operatorname{sign}(e_i)$, we know that ζ has a normal distribution with mean zero and covariance matrix $u_1^2\Sigma(1,\alpha)$.

As a consequence, $n^{\alpha/2}L^{-1/2}(n)(\hat{\theta}^{(1)} - \theta_0)$ converges in distribution to a piecewisely normal random vector v^*, where

$$v^* = \begin{cases} M_0\zeta \sim N(0, u_1^2 M_0 \Sigma(1,\alpha) M_0), & \text{if} \quad v^* \in S_0, \\ M_{i_1,\dots,i_k}\zeta \sim N(0, u_1^2 M_{i_1,\dots,i_k}\Sigma(1,\alpha)M_{i_1,\dots,i_k}), & \text{if} \quad v^* \in S_{i_1,\dots,i_k}, \ k = 1, 2, \dots. \end{cases}$$

3. Limiting distribution of L_2-estimate $\hat{\theta}^{(2)}$

In this section we consider the L_2-estimate $\hat{\theta}^{(2)}$ of parameter θ, that is the optimal solution of program (1.6). Here Assumption (C2) is changed as

(C2′) $\gamma(k) = k^{-\alpha}L(k)$, $0 < \alpha < 1/\tau_2$, where τ_2 is the Hermite rank of the function $G(x)$.

Let

$$a_{n2}^2 = n^{\tau_2\alpha}L^{-\tau_2}(n).$$

The following two theorems are similar to Theorems 2.1–2.2 in Section 2, so we omit the proofs.

THEOREM 3.1. *Suppose conditions* (C1), (C2′), (C3) *and* (C4) *are satisfied, then*

$$V_n(w) \to V(w) = w'\boldsymbol{K}w - 2w'\xi$$

in distribution uniformly for all w, *where* $\xi = (\xi_1, \dots, \xi_d)'$ *is a* d*-dimensional real-valued, mean zero random vector,*

$$\xi = \frac{l_{\tau_2}}{\tau_2!}D^{-\tau_2/2}\int_{R^{\tau_2}} K_0(x_1, \dots, x_{\tau_2})|x_1|^{(\alpha-1)/2}\cdots|x_{\tau_2}|^{(\alpha-1)/2}dW(x_1)\cdots dW(x_{\tau_2}),$$

here $l_{\tau_2} = E\{G(\eta_1)H_{\tau_2}(\eta_1)\}$ *and the covariance matrix of* ξ, $\frac{l_{\tau_2}^2}{\tau_2!}\Sigma(\tau_2,\alpha)$, *is positive definite with*

$$\Sigma(\tau_2,\alpha) = \lim_{n\to\infty} n^{\tau_2\alpha-1}L^{-\tau_2}(n)\sum_{i,j=1}^{n}\nabla f(x_i,\theta_0)(\nabla f(x_j,\theta_0))'(\gamma(i-j))^{\tau_2}.$$

THEOREM 3.2. *Under Assumptions* (C1), (C2′), (C3)–(C6), *we have*

(i) $n^{\tau_2\alpha/2}L^{-\tau_2/2}(n)(\hat{\theta}^{(2)} - \theta_0)$ *is bounded in probability;*

(ii) *If the set of optimal solutions of problem* (1.8) *is a singleton for each value of* ξ, *then* $n^{\tau_2\alpha/2}L^{-\tau_2/2}(n)(\hat{\theta}^{(2)} - \theta_0)$ *converges in distribution to the optimal solution* w^* *of problem* (1.8). *And*

$$w^* = \begin{cases} \widetilde{M_0}\xi, & \textit{if} \quad w^* \in S_0, \\ \widetilde{M}_{i_1,\dots,i_k}\xi, & \textit{if} \quad w^* \in S_{i_1,\dots,i_k}, \ k = 1, 2, \dots, \end{cases}$$

here $\widetilde{M_0}$, $\widetilde{M}_{i_1,\dots,i_k}$ *are the first blocks of the matrices* $\widetilde{H_0}^{-1}$, $\widetilde{H}_{i_1,\dots,i_k}^{-1}$ *respectively.* $\widetilde{H_0}$ *is defined by*

$$\widetilde{H_0} = \begin{pmatrix} \boldsymbol{K} & \nabla h \\ (\nabla h)' & 0 \end{pmatrix},$$

$\widetilde{H}_{i_1,\ldots,i_k}$ *is defined in the same way as* $H_{i_1,\ldots,i_k}$ *being replaced* $2F'(0)\boldsymbol{K}$ *by* $\boldsymbol{K}$.

Remark 3.1. As shown in Remark 2.1, for the case of $\tau_2 = 1$, ξ also has a normal distribution with mean zero and covariance matrix $l_1^2\Sigma(1,\alpha)$. Then $n^{\alpha/2}L^{-1/2}(n)(\hat{\theta}^{(2)} - \theta_0)$ converges in distribution to a piecewisely normal random vector w^*,

$$w^* = \begin{cases} \widetilde{M_0}\xi \sim N(0, l_1^2\widetilde{M_0}\Sigma(1,\alpha)\widetilde{M_0}), & \text{if} \quad w^* \in S_0, \\ \widetilde{M}_{i_1,\ldots,i_k}\zeta \sim N(0, l_1^2\widetilde{M}_{i_1,\ldots,i_k}\Sigma(1,\alpha)\widetilde{M}_{i_1,\ldots,i_k}), & \text{if} \quad w^* \in S_{i_1,\ldots,i_k}, \quad k = 1, 2, \ldots. \end{cases}$$

4. Asymptotic efficiencies of L_1-estimates and L_2-estimates

In this section we will list the asymptotic distributions of L_1-estimates and L_2-estimates in the i.i.d. errors case and in the long-range dependent errors case for unconstrained nonlinear regression. Then, together with the estimates derived in Section 2 and Section 3, we compare the asymptotic covariance structures of all these estimates.

It is well known that unconstrained L_1- and L_2-estimates with i.i.d. errors are asymptotically normal, and L_2-estimate is asymptotically more efficient than L_1-estimate under the i.i.d. Gaussian errors. For constrained case, Wang (1996) showed the convergence in distribution of L_2-estimates. Using similar arguments as that in Sections 2 and 3, we can even give the limiting distributions of L_1- and L_2-estimates.

THEOREM 4.1. *Suppose the errors* $\{e_i\}$ *are Gaussian, then under Assumptions* (C3)–(C6), *we have*

(i) $n^{1/2}(\hat{\theta}^{(1)} - \theta_0)$, $n^{1/2}(\hat{\theta}^{(2)} - \theta_0)$ *are bounded in probability;*

(ii) *If the sets of optimal solutions of problems* (1.7) *and* (1.8) *are singletons for each value of* ζ^* *and* ξ^*, *where* ζ^* *and* ξ^* *have the normal distribution* $N(0,\boldsymbol{K})$, *then* $n^{1/2}(\hat{\theta}^{(1)} - \theta_0)$, $n^{1/2}(\hat{\theta}^{(2)} - \theta_0)$ *converge in distribution to the optimal solution* v^* *and* w^* *of problems* (1.7) *and* (1.8), *respectively. And*

$$v^* = \begin{cases} M_0\zeta^*, & \text{if} \quad v^* \in S_0, \\ M_{i_1,\ldots,i_k}\zeta^*, & \text{if} \quad v^* \in S_{i_1,\ldots,i_k}, \quad k = 1, 2, \ldots, \end{cases}$$

$$w^* = \begin{cases} \widetilde{M_0}\xi^*, & \text{if} \quad w^* \in S_0, \\ \widetilde{M}_{i_1,\ldots,i_k}\xi^*, & \text{if} \quad w^* \in S_{i_1,\ldots,i_k}, \quad k = 1, 2, \ldots. \end{cases}$$

Due to the complexity of these limiting distributions, it is not easy to compare the asymptotic efficiencies of these estimators. However, if we know the location of v^* and w^*, then by Theorem 4.1 one can easily get the asymptotic covariance structure of the estimates.

For example, if v^*, $w^* \in S_1$, where

$$\begin{aligned} S_1 = \{v : (\nabla g_1(\theta_0))'v = 0; (\nabla g_i(\theta_0))'v < 0, i \in I \setminus \{1\}; \\ (\nabla h_j(\theta_0))'v = 0, j = 1, \ldots, p\}. \end{aligned}$$

Then we have

$$n^{1/2}(\hat{\theta}^{(1)} - \theta_0) \to N(0, M_1\boldsymbol{K}M_1),$$

and

$$n^{1/2}(\hat{\theta}^{(2)} - \theta_0) \to N(0, \widetilde{M_1}\boldsymbol{K}\widetilde{M_1})$$

in distribution. Elementary calculations yield

$$M_1 = (2F'(0)\boldsymbol{K})^{-1}[I - \nabla g_1((\nabla g_1)'\boldsymbol{K}^{-1}\nabla g_1)^{-1}(\nabla g_1)'\boldsymbol{K}^{-1} - \nabla h Q^{-1}(\nabla h)'\boldsymbol{K}^{-1} + R_1] \tag{4.1}$$

and

$$\widetilde{M_1} = \boldsymbol{K}^{-1}[I - \nabla g_1((\nabla g_1)'\boldsymbol{K}^{-1}\nabla g_1)^{-1}(\nabla g_1)'\boldsymbol{K}^{-1} - \nabla h Q^{-1}(\nabla h)'\boldsymbol{K}^{-1} + R_1], \tag{4.2}$$

where

$$\begin{aligned} R_1 &= ((\nabla g_1)'\boldsymbol{K}^{-1}\nabla g_1)^{-1}\nabla h Q^{-1}((\nabla h)'\boldsymbol{K}^{-1}\nabla g_1)(\nabla g_1)'\boldsymbol{K}^{-1} \\ &\quad + \nabla g_1((\nabla g_1)'\boldsymbol{K}^{-1}\nabla g_1)^{-1}((\nabla g_1)'\boldsymbol{K}^{-1}\nabla h)Q^{-1}(\nabla h)'\boldsymbol{K}^{-1} \\ &\quad - \nabla g_1((\nabla g_1)'\boldsymbol{K}^{-1}\nabla g_1)^{-1}((\nabla g_1)'\boldsymbol{K}^{-1}\nabla h)Q^{-1}((\nabla h)'\boldsymbol{K}^{-1}\nabla g_1) \\ &\quad \times (\nabla g_1)'\boldsymbol{K}^{-1}((\nabla g_1)'\boldsymbol{K}^{-1}\nabla g_1)^{-1}, \\ Q &= (\nabla h)'\boldsymbol{K}^{-1}\nabla h - ((\nabla h)'\boldsymbol{K}^{-1}\nabla g_1)((\nabla g_1)'\boldsymbol{K}^{-1}\nabla g_1)^{-1}((\nabla g_1)'\boldsymbol{K}^{-1}\nabla h). \end{aligned}$$

Since $2F'(0) = \sqrt{\frac{2}{\pi}}$, this implies that

$$M_1\boldsymbol{K}M_1 = \frac{\pi}{2}\widetilde{M_1}\boldsymbol{K}\widetilde{M_1}.$$

Then we can conclude that L_2-estimate is asymptotically more efficient than L_1-estimate for constrained model with the i.i.d. Gaussian errors.

Next let us concentrate on the long-range dependent errors case for the unconstrained model. We note that the case of $v^* \in S_0$ or $w^* \in S_0$ is essentially equivalent to an unconstrained problem, if we don't consider the equality constraints $h_j(\theta) = 0$, $j = 1, \ldots, p$. Thus it follows from Theorems 2.2 and 3.2 that the limiting distribution of L_1-estimates, $n^{\tau_1\alpha/2}L^{-\tau_1/2}(n)(\hat{\theta}^{(1)} - \theta_0)$, is equal to $(2F'(0)\boldsymbol{K})^{-1}\zeta$ and L_2-estimates, $n^{\tau_2\alpha/2}L^{-\tau_2/2}(n)(\hat{\theta}^{(2)} - \theta_0)$, converges in distribution to $\boldsymbol{K}^{-1}\xi$.

If $G(x) = x$, i.e. the errors $\{e_i\}$ are Gaussian, it is easy to see that $\tau_1 = \tau_2 = 1$, $u_1 = \sqrt{\frac{2}{\pi}}$, $l_1 = 1$ and $2F'(0) = \sqrt{\frac{2}{\pi}}$, then from Remarks 2.1 and 3.1, we know that $\zeta \sim N(0, \frac{2}{\pi}\Sigma(1,\alpha))$ and $\xi \sim N(0, \Sigma(1,\alpha))$. Hence $n^{\alpha/2}L^{-1/2}(n)(\hat{\theta}^{(1)} - \theta_0)$ and $n^{\alpha/2}L^{-1/2}(n)(\hat{\theta}^{(2)} - \theta_0)$ have the same normal limiting distribution $N(0, \boldsymbol{K}^{-1}\Sigma(1,\alpha)\boldsymbol{K}^{-1})$. This implies that, if the long-range dependent errors are Gaussian, L_1-estimates and L_2-estimates will have the same asymptotic covariance structure for unconstrained nonlinear regressions, which coincide with the results in Koul and Murkherjee (1993) and Koul (1996).

If there are restrictions of parameters appeared in nonlinear regression models, we still have the asymptotic equivalence of L_1-estimates and L_2-estimates for the Gaussian errors case. The asymptotic equivalence is not affected by the appearance of constraints.

For instance, again suppose the errors $\{e_i\}$ are Gaussian and v^*, $w^* \in S_1$, we have

$$n^{\alpha/2}L^{-1/2}(n)(\hat{\theta}^{(1)} - \theta_0) \to N(0, u_1^2 M_1\Sigma(1,\alpha)M_1),$$

and

$$n^{\alpha/2}L^{-1/2}(n)(\hat{\theta}^{(2)}-\theta_0) \to N(0, l_1^2\widetilde{M_1}\Sigma(1,\alpha)\widetilde{M_1}).$$

By (4.1), (4.2) and recall that $u_1^2 = \frac{2}{\pi}l_1^2$ and $2F'(0) = \sqrt{\frac{2}{\pi}}$, we come to the result that

$$u_1^2 M_1\Sigma(1,\alpha)M_1 = l_1^2\widetilde{M_1}\Sigma(1,\alpha)\widetilde{M_1}.$$

Finally, compared to L_2-estimates, L_1-estimator is asymptotically more efficient at the double exponential error distribution, the logistic error distribution. Let us briefly explain this conclusion. For convenience again suppose v^*, $w^* \in S_1$, by Theorems 2.2 and 3.2 we see that

$$n^{\tau_1\alpha/2}L^{-\tau_1/2}(n)(\hat{\theta}^{(1)}-\theta_0) \to M_1\zeta \tag{4.3}$$

and

$$n^{\tau_2\alpha/2}L^{-\tau_2/2}(n)(\hat{\theta}^{(2)}-\theta_0) \to \widetilde{M_1}\xi. \tag{4.4}$$

Moreover, from Theorems 2.1 and 3.1, we know that

$$\mathrm{Var}(\xi) = \frac{l_{\tau_2}^2}{\tau_2!}\Sigma(\tau_2,\alpha), \qquad \mathrm{Var}(\zeta) = \frac{u_{\tau_1}^2}{\tau_1!}\Sigma(\tau_1,\alpha). \tag{4.5}$$

Example 4.1. Suppose we have the logistic error distribution, i.e. $F(x) = (1+e^{-x})^{-1}$. Clearly this distribution satisfies the Assumption (C1) and $F'(0) = 1/4$ (so we have $M_1 = 2\widetilde{M_1}$), $G(x) = F^{-1}(\Phi(x)) = \ln(\frac{\Phi(x)}{1-\Phi(x)})$. Therefore

$$u_1 = E\{\mathrm{sign}(G(\eta_1))\eta_1\} = 2\int_0^\infty x\phi(x)dx = \sqrt{\frac{2}{\pi}} \neq 0$$

and

$$\begin{aligned} l_1 &= E\{G(\eta_1)\eta_1\} = 2\int_0^\infty \ln\left(\frac{\Phi(x)}{1-\Phi(x)}\right)x\phi(x)dx \\ &= 2\int_{-\infty}^\infty \ln(\Phi(x))x\phi(x)dx = 2\int_{-\infty}^\infty \phi^2(x)/\Phi(x)dx \\ &> \frac{1}{\pi}\int_C^a e^{-x^2}/\Phi(x)dx \geq \frac{a-C}{\pi}e^{-a^2}/\Phi(a) \\ &\geq 2u_1 > 0, \end{aligned}$$

where a, C $(a > C)$ are negative constants such that $\frac{a-C}{\pi}e^{-a^2}/\Phi(a) \geq 2\sqrt{\frac{2}{\pi}}$. This means that $\tau_1 = \tau_2 = 1$ and $l_1^2 \geq 4u_1^2$. Then by (4.1)–(4.5), we conclude that for logistic error distribution the asymptotic covariance of $n^{\tau_2\alpha/2}L^{-\tau_2/2}(n)(\hat{\theta}^{(2)}-\theta_0)$ is much larger than the asymptotic covariance of $n^{\tau_1\alpha/2}L^{-\tau_1/2}(n)(\hat{\theta}^{(1)}-\theta_0)$. This implies that L_1-estimator is more efficient than L_2-estimator at logistic error distribution in constrained nonlinear regression with long-range dependence.

Example 4.2. Assume the d.f. F of the errors has a density function $f(x) = \frac{1}{2}e^{-|x|}$. This double exponential distribution satisfies the Assumption (C1) and $F'(0) = 1/2$ (we have $M_1 = \widetilde{M_1}$),

$$G(x) = \begin{cases} -\ln(2\Phi(-x)), & x \geq 0, \\ \ln(2\Phi(x)), & x < 0. \end{cases}$$

Hence

$$u_1 = E\{\mathrm{sign}(G(\eta_1))\eta_1\} = 2\int_0^\infty x\phi(x)dx = \sqrt{\frac{2}{\pi}} \neq 0$$

and

$$\begin{aligned} l_1 &= E\{G(\eta_1)\eta_1\} = 2\int_{-\infty}^0 \ln(2\Phi(x))x\phi(x)dx \\ &= -2\int_{-\infty}^0 \ln(2\Phi(x))d\phi(x) = \int_{-\infty}^0 \phi^2(x)/\Phi(x)dx \\ &> \frac{1}{2\pi}\int_C^a e^{-x^2}/\Phi(x)dx \geq \frac{a-C}{2\pi}e^{-a^2}/\Phi(a) \\ &\geq u_1 > 0, \end{aligned}$$

where a, C ($a > C$) are the same as that in Example 4.1. Then by (4.1)–(4.5), we know that L_1-estimator is more efficient than L_2-estimator at double exponential error distribution in constrained nonlinear regression with long-range dependence.

Acknowledgements

This paper is partly supported by NSFC and NSFJS. The author expresses her sincere gratitude for the referees' helpful comments and suggestions that improved the early version of the paper. The author is also grateful to Professors M. Csörgő and B. Szyszkowicz for their hospitality during her stay in Carleton University's School of Mathematics & Statistics.

References

Aitchison, J. and Silvey, S. V. (1958). Maximum likelihood estimation of parameters subject to restraints, *The Annals of Mathematical Statistics*, **29**, 813–828.

Beran, J. (1992). Statistical methods for data with long-range dependence, *Statistical Science*, **7**, 404–427.

Dahlhaus, R. (1995). Efficient location and regression estimation for long range dependent regression models, *The Annals of Statistics*, **23**, 1029–1047.

Dobrushin, R. L. and Major, P. (1979). Noncentral limit theorems for nonlinear functionals of Gaussian fields, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **50**, 27–52.

Giraitis, L., Koul, H. L. and Surgailis, D. (1996). Asymptotic normality of regression estimators with long memory errors, *Statistics & Probability Letters*, **29**, 317–335.

Huber, P. (1967). The behavior of maximum likelihood estimates under nonstandard conditions, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 221–233, University of California Press.

Joreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis, *Psychometrika*, **34**, 183–202.

Koul, H. L. (1996). Asymptotics of M-estimators in nonlinear regression with long-range dependent errors, *Athens Conference on Applied Probability and Time Series Analysis*, Vol. II, Lecture Notes in Statistics, Vol. 115, 272–290, Springer, New York.

Koul, H. L. and Mukherjee, K. (1993). Asymptotics of R-, MD- and LAD-estimators in linear regression models with long range dependent errors, *Probability Theory and Related Fields*, **95**, 535–553.

Liew, C. K. (1976). Inequality constrained least squares estimations, *Journal of the American Statistical Association*, **71**, 746–751.

Nagaraj, N. and Fuller, W. (1991). Estimation of the parameters of linear time series models subject to nonlinear regression, *The Annals of Statistics*, **19**, 1143–1154.

Robinson, P. M. (1995). Log-periodogram regression of time series with long range dependence, *The Annals of Statistics*, **23**, 1048–1072.

Robinson, P. M. and Hidalgo, F. J. (1997). Time series regression with long range dependence, *The Annals of Statistics*, **25**, 77–104.

Rockafellar, R. T. (1970). *Convex Analysis*, Princeton University Press, Princeton, New Jersey.

Sen, P. K. (1979). Asymptotic properties of maximum likelihood estimators based on conditional specification, *The Annals of Statistics*, **7**, 1019–1033.

Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis, *International Statistical Review*, **56**, 49–62.

Silvey, S. D. (1959). The Lagrangian multiplier test, *The Annals of Mathematical Statistics*, **30**, 389–407.

Taqqu, M. S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **31**, 287–302.

Taqqu, M. S. (1979). Convergence of integrated processes of arbitrary Hermite rank, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **50**, 53–83.

Wang, J. (1996). Asymptotics of least-squares estimators for constrained nonlinear regression, *The Annals of Statistics*, **24**, 1316–1326.

Yajima, Y. (1988). On estimation of a regression model with long-memory stationary errors, *The Annals of Statistics*, **16**, 791–807.

Yajima, Y. (1991). Asymptotic properties of the LSE in a regression model with long-memory stationary erors, *The Annals of Statistics*, **19**, 158–177.