isid/ms/2003/05
January 22, 2003
http://www.isid.ac.in/~statmath/eprints

Estimation of cusp in nonregular nonlinear regression models

B. L. S. Prakasa Rao

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

ESTIMATION OF CUSP IN NONREGULAR NONLINEAR REGRESSION MODELS

B.L.S. PRAKASA RAO
INDIAN STATISTICAL INSTITUTE,NEW DELHI

Abstract

The asymptotic properties of the least squares estimator of the cusp in some nonlinear nonregular regression models is investigated via the study of the weak convergence of the least squares process generalizing earlier results in Prakasa Rao (*Statist. Prob. Lett.* **3** (1985), 15-18).

Key words: Cusp, Least squares process, Nonlinear regression, Nonregular, Fractional Brownian motion.

AMS Subject classification (2000): Primary 62 J 02.

1 Introduction

Consider the nonlinear regression model

$$Y_i = S(x_i, \theta) + \varepsilon_i, i \geq 1 \tag{1. 1}$$

We now discuss the problem of estimation of the parameter θ by the least squares approach when the parameter θ is a cusp for the function $S(x,\theta) = s(x-\theta)$. This problem is not amenable to standard methods using the Taylor's expansion as the function $S(x,\theta)$ is not differentiable at θ. We study the asymptotic properties of the least squares estimator via the least squares process developed earlier in Prakasa Rao (1986) (cf. Prakasa Rao (1987)). A special case of the problem is studied in Prakasa Rao (1985).

2 Main Result

Consider the nonlinear regression model

$$Y_i = S(x_i, \theta) + \varepsilon_i, i \geq 1 \tag{2. 1}$$

where

$$\begin{aligned} S(x,\theta) &= a|x-\theta|^{\lambda} + h(x-\theta), x \leq \theta \\ S(x,\theta) &= b|x-\theta|^{\lambda} + h(x-\theta), x \geq \theta \end{aligned} \tag{2. 2}$$

where $a \neq 0, b \neq 0, 0 < \lambda < \frac{1}{2}$ and $h(.)$ satisfies the Holder condition of order $\beta > \lambda + \frac{1}{2}$. Further more $\{\varepsilon_i, i \geq 1\}$ are independent and identically distributed random variables with mean zero and known finite positive variance σ^2 (say). Without loss of generality, we assume that $\sigma^2 = 1$. Further suppose that $\theta \in \Theta$ compact contained in R.

Let θ_0 denote the true parameter. Suppose $\{x_i, i \geq 1\}$ is a real sequence with the property that

$$\sum_{i=1}^{n} \{S(x_i,\theta) - S(x_i,\theta_0)\}^2 = 2nC(\lambda)|\theta - \theta_0|^{2\lambda+1}(1 + o(1)) \tag{2. 3}$$

as $n \to \infty$ where $C(\lambda) \neq 0$ and further suppose that there exists $0 < k_1 < k_2 < \infty$ such that

$$nk_1|\theta_1 - \theta_2|^{2\lambda+1} \leq \sum_{i=1}^{n} \{S(x_i,\theta_1) - S(x_i,\theta_2)\}^2 \leq nk_2|\theta_1 - \theta_2|^{2\lambda+1} \tag{2. 4}$$

for all θ_1 and θ_2 in Θ.

Let $\hat{\theta}_n$ be a least squares estimator (LSE) of θ obtained by minimizing

$$Q_n(\theta) = \sum_{i=1}^{n} (Y_i - S(x_i,\theta))^2. \tag{2. 5}$$

It is obvious that $\hat{\theta}_n$ minimizes

$$
\begin{aligned}
(2.\ 6)\quad Q_n(\theta) - Q_n(\theta_0) &= \sum_{i=1}^{n}(Y_i - S(x_i,\theta))^2 - \sum_{i=1}^{n}(Y_i - S(x_i,\theta_0))^2 \\
&= 2\sum_{i=1}^{n}\varepsilon_i(S(x_i,\theta) - S(x_i,\theta_0)) + \sum_{i=1}^{n}(S(x_i,\theta) - S(x_i,\theta_0))^2.
\end{aligned}
$$

Observe that

$$
\begin{aligned}
(2.\ 7)\qquad E_{\theta_0}[Q_n(\theta) - Q_n(\theta_0)] &= \sum_{i=1}^{n}(S(x_i,\theta) - S(x_i,\theta_0))^2 \\
&= 2nC(\lambda)|\theta - \theta_0|^{2\lambda+1}(1 + o(1))
\end{aligned}
$$

and

$$
\begin{aligned}
(2.\ 8)\qquad Var_{\theta_0}[Q_n(\theta) - Q_n(\theta_0)] &= 4\sum_{i=1}^{n}(S(x_i,\theta) - S(x_i,\theta_0))^2 \\
&= 8nC(\lambda)|\theta - \theta_0|^{2\lambda+1}(1 + o(1)).
\end{aligned}
$$

In general there exists $k_2 > 0$ independent of n, θ and $\theta_0 \in \Theta$ such that

$$
(2.\ 9)\qquad E_{\theta_0}[Q_n(\theta) - Q_n(\theta_0)] \leq nk_2|\theta - \theta_0|^{2\lambda+1}
$$

and

$$
(2.\ 10)\qquad Var_{\theta_0}[Q_n(\theta) - Q_n(\theta_0)] \leq 4nk_2|\theta - \theta_0|^{2\lambda+1}.
$$

Further more

$$
\begin{aligned}
(2.\ 11)\qquad & Cov_{\theta_0}[Q_n(\theta_1) - Q_n(\theta_0), Q_n(\theta_2) - Q_n(\theta_0)] \\
&= 4\sum_{i=1}^{n}(S(x_i,\theta_1) - S(x_i,\theta_0))(S(x_i,\theta_2) - S(x_i,\theta_0)) = \\
&= 4nC(\lambda)[|\theta_1 - \theta_0|^{2\lambda+1} + |\theta_2 - \theta_0|^{2\lambda+1} - |\theta_1 - \theta_2|^{2\lambda+1}](1 + o(1))
\end{aligned}
$$

from the relation

$$||f||^2 + ||g||^2 - ||f - g||^2 = 2 < f, g >$$

for any two vectors f and g in R^n. Let

$$J_n(\theta) = Q_n(\theta) - Q_n(\theta_0).$$

In view of the above relations , it follows by arguments similar to those given in the Theorem 3.6 of Prakasa Rao (1968a) that there exists $\eta > 0$ such that

$$\lim_{\tau\to\infty}\limsup_{n\to\infty} P_{\theta_0}\Big[\inf_{|\theta-\theta_0|>\tau n^{-\rho}} \frac{J_n(\theta)}{n|\theta-\theta_0|^{2\lambda+1}} \leq \eta\Big] = 0 \tag{2. 12}$$

where $\rho = (2\lambda+1)^{-1}$. In fact the same proof shows that , for any $\tau > 0$,

$$P_{\theta_0}[n^\rho|\hat{\theta}_n - \theta_0| > \tau] \leq C\tau^{-(2\lambda+1)} \tag{2. 13}$$

where the constant C is independent of n and τ. In view of the above observation , the process $\{J_n(\theta), \theta \in \Theta\}$ has a minimum in the interval $[\theta_0 - \tau n^{-\rho}, \theta_0 + \tau n^{-\rho}]$ with probability approaching one for large τ. For any such $\tau > 0$, let

$$R_n(\phi) = J_n(\theta_0 + \phi n^{-\rho}), \phi \in [-\tau, \tau] \tag{2. 14}$$

and let $R(\phi)$ be a gaussian process on $[-\tau, \tau]$ with

$$E[R(\phi)] = 2C(\lambda)|\phi|^{2\lambda+1} \tag{2. 15}$$

and

$$Cov[R(\phi_1), R(\phi_2)] = 4C(\lambda)[|\phi_1|^{2\lambda+1} + |\phi_2|^{2\lambda+1} - |\phi_1 - \phi_2|^{2\lambda+1}]. \tag{2. 16}$$

Observe that the process $\{\tilde{R}(\phi), -\infty < \phi < \infty\}$ where

$$\tilde{R}(\phi) = \frac{R(\phi) - E[R(\phi)]}{\sqrt{8C(\lambda)}} \tag{2. 17}$$

is a gaussian random process with mean zero and the covariance function

$$Cov(\tilde{R}(\phi_1), \tilde{R}(\phi_2)) = \frac{1}{2}[|\phi_1|^{2\lambda+1} + |\phi_2|^{2\lambda+1} - |\phi_1 - \phi_2|^{2\lambda+1}] \tag{2. 18}$$

which is the fractional Brownian motion with the Hurst parameter $H = \lambda + \frac{1}{2}$.

Let

$$\begin{aligned} Z_n(\phi) &= \sum_{i=1}^n \varepsilon_i(S(x_i, \theta_0 + \phi n^{-\rho}) - S(x_i, \theta_0)) \\ &= \sum_{i=1}^n a_{ni}\varepsilon_i \end{aligned} \tag{2. 19}$$

and

$$T_n(\phi) = \sum_{i=1}^n (S(x_i, \theta_0 + \phi n^{-\rho}) - S(x_i, \theta_0))^2. \tag{2. 20}$$

Observe that $T_n(\phi)$ is continuous in ϕ for any fixed $n \geq 1$ and

$$T_n(\phi) \to 2C(\lambda)|\phi|^{2\lambda+1} \text{ as } n \to \infty. \tag{2. 21}$$

In addition

$$(2.\ 22) \qquad Z_n(\phi) \xrightarrow{\mathcal{L}} N(0, 2C(\lambda)|\phi|^{2\lambda+1}) \text{ as } n \to \infty$$

since $\{\varepsilon_i, i \geq 1\}$ are independent and identically distributed random variables with mean zero and finite positive variance and $\{a_{nk}, 1 \leq k \leq n\}$ satisfy the condition

$$(2.\ 23) \qquad \max_{1 \leq k \leq n} \frac{a_{nk}^2}{\sum_{i=1}^n a_{ni}^2} \to 0 \text{ as } n \to \infty.$$

This follows from a central limit theorem due to Eicker (1963). The above relation can be proved by the following arguments. Note that

$$(2.\ 24) \qquad \begin{aligned} \sum_{i=1}^n a_{ni}^2 &= \sum_{i=1}^n (S(x_i, \theta_0 + \phi n^{-\rho}) - S(x_i, \theta_0))^2 \\ &= T_n(\phi) \end{aligned}$$

which tends to $2C(\lambda)|\phi|^{2\lambda+1}$ as $n \to \infty$ by the relation (2.21) and the latter in turn implies that

$$\max_{1 \leq i \leq n} (S(x_i, \theta_0 + \phi n^{-\rho}) - S(x_i, \theta_0))^2 \to 0.$$

This follows from the observation that if $\sum_{1 \leq k \leq n} a_{nk}^2 \to c > 0$ and $\max_{1 \leq k \leq N} a_{nk}^2 \to 0$ for every fixed $N \geq 1$, then $\max_{1 \leq k \leq n} a_{nk}^2 \to 0$ as $n \to \infty$. The above discussion proves (2.22). Similarly it can be shown that all the finite dimensional distributions of the process $\{Z_n(\phi), -\tau \leq \phi \leq \tau\}$ converge to the corresponding finite dimensional distributions of the gaussian process $\{Z(\phi), -\tau \leq \phi \leq \tau\}$with mean zero and

$$(2.\ 25) \qquad Cov[Z(\phi_1), Z(\phi_2)] = C(\lambda)[|\phi_1|^{2\lambda+1} + |\phi_2|^{2\lambda+1} - |\phi_1 - \phi_2|^{2\lambda+1}].$$

In addition, observe that

$$(2.\ 26) \qquad \begin{aligned} E|Z_n(\phi_1) - Z_n(\phi_2)|^2 &= \sum_{i=1}^n (S(x_i, \theta_0 + \phi_1 n^{-\rho}) - S(x_i, \theta_0 + \phi_2 n^{-\rho}))^2 \\ &\leq k_2 |\phi_1 - \phi_2|^{2\lambda+1}. \end{aligned}$$

where k_2 is independent of n, ϕ_1 and ϕ_2. Hence the family of measures $\{\mu_n\}$ generated by the stochastic processes $\{Z_n(\phi), -\tau \leq \phi \leq \tau\}$ on the space $C[-\tau, \tau]$ of continuous functions on the interval $[\tau, \tau]$ with supremum norm topology forms a tight family. This observation together with the fact that the finite dimensional distributions of the process $\{Z_n(\phi), -\tau \leq \phi \leq \tau\}$ converge weakly to the corresponding finite dimensional distributions of the process $\{Z(\phi), -\tau \leq \phi \leq \tau\}$ prove that the sequence of processes $\{Z_n(\phi), -\tau \leq \phi \leq \tau\}$ converge weakly to the gaussian process $\{Z(\phi), -\tau \leq \phi \leq \tau\}$. Hence the sequence of processes $\{R_n(\phi), -\tau \leq \phi \leq \tau\}$ converge weakly to the gaussian process $\{R(\phi), -\tau \leq \phi \leq \tau\}$ with mean function and covariance function given by (2.15) and (2.16) respectively. It now follows by arguments similar to those given in Prakasa Rao (1968a) that

(2. 27) $$n^{\frac{1}{2\lambda+1}}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \hat{\phi} \text{ as } n \to \infty$$

where $\hat{\phi}$ has the distribution of the location of the minimum of the gaussian process $\{R(\phi), -\infty < \phi < \infty\}$ with

(2. 28) $$E[R(\phi)] = 2C(\lambda)|\phi|^{2\lambda+1}$$

and

(2. 29) $$Cov[R(\phi_1), R(\phi_2)] = 4C(\lambda)[|\phi_1|^{2\lambda+1} + |\phi_2|^{2\lambda+1} - |\phi_1 - \phi_2|^{2\lambda+1}].$$

We now have the following main result.

Theorem 2.1: Consider the nonlinear regression model

(2. 30) $$Y_i = S(x_i, \theta) + \varepsilon_i, i \geq 1$$

where

(2. 31) $$S(x, \theta) = a|x - \theta|^\lambda + h(x - \theta), x \leq \theta$$
$$S(x, \theta) = b|x - \theta|^\lambda + h(x - \theta), x \geq \theta$$

where $a \neq 0, b \neq 0, 0 < \lambda < \frac{1}{2}, \theta \in \Theta$ and $h(.)$ satisfies the Holder condition of order $\beta > \lambda + \frac{1}{2}$. Suppose Θ is compact. Let θ_0 be the true parameter. Further more suppose that $\{\varepsilon_i, i \geq 1\}$ are independent and identically distributed random variables with mean zero and variance one. Let $\{x_i\}$ be a real sequence satisfying

(2. 32) $$\sum_{i=1}^{n}\{S(x_i, \theta) - S(x_i, \theta_0)\}^2 = 2nC(\lambda)|\theta - \theta_0|^{2\lambda+1}(1 + o(1))$$

as $n \to \infty$ where $C(\lambda) \neq 0$ and further suppose that there exists $0 < k_1 < k_2 < \infty$ such that

(2. 33) $$nk_1|\theta_1 - \theta_2|^{2\lambda+1} \leq \sum_{i=1}^{n}\{S(x_i, \theta_1) - S(x_i - \theta_2)\}^2 \leq nk_2|\theta_1 - \theta_2|^{2\lambda+1}$$

for all θ_1 and θ_2 in Θ. Let $\hat{\theta}_n$ be a least squares estimator (LSE) of θ based on the observations $\{Y_i, 1 \leq i \leq n\}$. Then there exists a constant $C > 0$ such that for any $\tau > 0$ and for any $n \geq 1$,

(2. 34) $$P_{\theta_0}[n^\rho|\hat{\theta}_n - \theta_0| > \tau] \leq C\tau^{-(2\lambda+1)}$$

and

(2. 35) $$n^\rho(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \hat{\phi} \text{ as } n \to \infty$$

where $\hat{\phi}$ is the location of the minimum of the process $\{R(\phi), -\infty < \phi < \infty\}$ with

$$E[R(\phi)] = 2C(\lambda)|\phi|^{2\lambda+1} \tag{2. 36}$$

and

$$Cov[R(\phi_1), R(\phi_2)] = 4C(\lambda)[|\phi_1|^{2\lambda+1} + |\phi_2|^{2\lambda+1} - |\phi_1 - \phi_2|^{2\lambda+1}]. \tag{2. 37}$$

Here $\rho = \frac{1}{2\lambda+1}$. The process $\{R(\phi), -\infty < \phi < \infty\}$ can be represented in the form

$$R(\phi) = \sqrt{8C(\lambda)}W^H(\phi) + 2C(\lambda)|\phi|^{2\lambda+1}$$

where W^H is the fractional Brownian motion with mean zero and the covariance function

$$Cov(W^H(\phi_1), W^H(\phi_2)) = \frac{1}{2}[|\phi_1|^{2\lambda+1} + |\phi_2|^{2\lambda+1} - |\phi_1 - \phi_2|^{2\lambda+1}].$$

Here $H = 2\lambda + 1$ is the Hurst parameter.

3 Remarks:

(i) Observe that $\rho > \frac{1}{2}$ if $0 < \lambda < \frac{1}{2}$ and the asymptotic variance is of the order $O(n^{-2\rho})$ which is small compared to the case when the regression function $S(x, \theta)$ is smooth and the asymptotic variance is of the order $O(n^{-1})$ (cf. Prakasa Rao (1984)). Let

$$\begin{aligned} d(x) &= a \text{ if } x < 0 \\ &= b \text{ if } x > 0. \end{aligned}$$

The condition (2.3) on the sequence $\{x_i\}$is a plausible condition that can be assumed. This can be justified by the following arguments.

Suppose the sequence $\{x_i, i \geq 1\}$ is the realization of a sequence of independent identically distributed random variables $\{X_i, i \geq 1\}$ with probability density function

$$\begin{aligned} f(x, \theta) &= h(x - \theta) \exp\{a(x - \theta)|x - \theta|^\lambda\} \text{ if } x \leq \theta \\ &= h(x - \theta) \exp\{b(x - \theta)|x - \theta|^\lambda\} \text{ if } x \geq \theta. \end{aligned}$$

The the Strong law of large numbers implies that

$$n^{-1} \sum_{i=1}^{n} \{S(X_i, \theta_1) - S(X_i, \theta_0)\}\{S(X_i, \theta_2) - S(X_i, \theta_0)\}$$

converges almost surely to

$$E[\{S(X_1, \theta_1) - S(X_1, \theta_0)\}\{S(X_1, \theta_2) - S(X_i, \theta_0)\}]$$

which is equal to

$$C(\lambda)[|\theta_1|^{2\lambda+1} + |\theta_2|^{2\lambda+1} - |\theta_1 - \theta_2|^{2\lambda+1}].$$

This can be seen from the Lemma 4 in Dachian and Kutoyants (2002). Infact

(3. 1)
$$\int_{-\infty}^{\infty} [d(x-\theta_1)|x-\theta_1|^p - d(x-\theta)|x-\theta|^\lambda][d(x-\theta_2)|x-\theta_2|^p - d(x-\theta)|x-\theta|^\lambda]dx =$$
$$= C(\lambda)[|\theta_1|^{2\lambda+1} + |\theta_2|^{2\lambda+1} - |\theta_1 - \theta_2|^{2\lambda+1}]$$

where

$$C(\lambda) = \frac{\Gamma(2\lambda+1)\Gamma(\frac{1}{2}-\lambda)}{2^{2\lambda+1}\sqrt{\pi}(2\lambda+1)}[a^2 + b^2 - 2ab\ \ cos(\pi\lambda)].$$

A special case of this result, when $a = b$, can be seen in Prakasa Rao (1968b). For the general case, see Ibragimov and Khasminskii (1981).

(ii) If $\lambda = \frac{1}{2}$ in the model (2.1), and the conditions stated in (2.3) and (2.4) hold, then it can be checked by similar arguments as before that there exists a constant $C > 0$ such that

(3. 2) $$P_{\theta_0}(n^{1/2}|\hat{\theta}_n - \theta_0| > \tau) \leq C\tau^{-2}$$

and

(3. 3) $$n^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \hat{\phi} \text{ as } n \to \infty$$

where $\hat{\phi}$ is the location of the minimum of the gaussian process $\{R(\phi), -\infty < \phi < \infty\}$ with mean function $E(R(\phi)) = 2C^*\phi^2$ and $Cov(R(\phi_1), R(\phi_2)) = 8C^*\phi_1\phi_2$ for some constant $C^* > 0$. It is easy to see that the process $R(\phi)$ can be represented in the form

(3. 4) $$R(\phi) = 2C^*\phi^2 + L\phi\psi$$

where $L = \sqrt{8C^*}$ and ψ is a standard normal random variable. Hence

(3. 5) $$\hat{\phi} = -\frac{L}{4C^*}\psi.$$

Combining the above remarks, we obtain that

(3. 6) $$(i)\hat{\theta}_n \xrightarrow{p} \theta_0 \text{ as } n \to \infty,$$

there exists a constant $C > 0$ independent of n and τ such that

(3. 7) $$(ii)P_{\theta_0}(n^{1/2}|\hat{\theta}_n - \theta_0| > \tau) \leq C\tau^{-2}$$

and

(3. 8) $$(iii)n^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N(0, (2C^*)^{-1}) \text{ as } n \to \infty.$$

Note that the limiting distribution of the least squares estimator is normal if $\lambda = \frac{1}{2}$ in the model . This case illustrates the situation when the standard regularity conditions do not hold and yet the estimator is strongly consistent and asymptotically normal.

References

Dachian, S., and Kutoyants, Y. (2002) On cusp estimation of ergodic diffusion process, Preprint.

Eicker, F. (1963) Central limit theorems for families of sequences of random variables, *Ann. Math. Statist.* **34**, 439-466.

Ibragimov, I.A. and Khasminskii, R.Z. (1981) *Statistical Estimation: asymptotic Theory*, springer-Verlag, New York.

Prakasa Rao, B.L.S. (1968a) Estimation of the location of the cusp of a continuous density, *Ann. Math. Statist.* **39**, 76-87.

Prakasa Rao, B.L.S. (1968b) On evaluating a certain integral, *Amer. Math. Monthly* **75**

Prakasa Rao, B.L.S. (1984) On the rate of convergence of the least squares estimator in the nonlinear regression model with dependent errors, *J. Multvariate Anal.* **14**, 315-322.

Prakasa Rao, B.L.S. (1985) Asymptotic theory of least squares estimator in a nonregular nonlinear regression model, *Statist. Prob. Lett.* **3**, 15-18.

Prakasa Rao, B.L.S. (1986) Weak convergence of the least squares process in the smooth case, *Statistics* **17**, 505-516.

Prakasa Rao, B.L.S. (1987) *Asymptotic Theory of Statistical Inference*, Wiley, New York.