Estimation of cusp in nonregular nonlinear regression models

B.L.S. Prakasa Rao

Indian Statistical Institute, 7, SJS Sansanwal Marg, New Delhi, 110016, India

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. Probab. Lett. 3 (1985) 15).

Keywords: Cusp; Least squares process; Nonlinear regression; Nonregular problem; Fractional Brownian motion

1. Introduction

Nonlinear regression models are widely used for modeling of stochastic phenomena. Several examples for such modeling are given by Bard [2]. The study of asymptotic properties of the least squares estimator (LSE) of the parameters occurring in nonlinear regression models has been the subject of investigation since it is in general difficult to obtain the exact distribution of the LSE for any fixed sample. Jennrich [9], Malinvaud [10], Bunke and Schmidt [3] and Wu [20], among others, have investigated the asymptotic properties of the LSE for nonlinear regression models. All the works cited above, on the asymptotic distribution theory for the LSE in nonlinear regression models, assume regularity conditions which include in particular the condition on the twice differentiability of the regression function with respect to the parameter in addition to other conditions (cf. [7]). We have given an alternate approach to the problem which circumvents the condition on the differentiability of the regression function in a series of papers in Prakasa Rao [15–17]. A comprehensive presentation of this approach is given in Prakasa Rao [18]. A review of these and related results is given in Prakasa Rao [19]. Detailed discussion of this approach for a special class of nonregular nonlinear regression models is given in Prakasa Rao [14]. Maximum likelihood estimation for cusp-type nonregular problems was first started in Prakasa Rao [11] and later by Ibragimov and Khasminskii [8]. Other works dealing with problems of estimation for nonregular models include Akahira and Takeuchi [1], Dachian [4] for the estimation of the cusp of Poisson observations and Dachian and Kutoyants [5] on the cusp estimation for ergodic diffusion process among others.

Consider the nonlinear regression model

$$Y_i = S(x_i, \theta) + \varepsilon_i, \quad i \geqslant 1. \tag{1.1}$$

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 in general is not amenable to standard methods using the Taylor's expansion, for instance, the function $S(x,\theta) = |x-\theta|^\lambda,\ 0<\lambda<\frac{1}{2}$ is not differentiable at θ. We study the asymptotic properties of the LSE via the least squares process developed in Prakasa Rao [15] (cf. [18,19]). A special case of this problem was studied in Prakasa Rao [14].

2. Main result

We now have the following main result.

Theorem 2.1. *Consider the nonlinear regression model*

$$Y_i = S(x_i, \theta) + \varepsilon_i, \quad i \geqslant 1, \tag{2.1}$$

where

$$\begin{aligned} S(x,\theta) &= a|x-\theta|^\lambda + h(x-\theta), \quad x \leqslant \theta, \\ S(x,\theta) &= b|x-\theta|^\lambda + h(x-\theta), \quad x \geqslant \theta, \end{aligned} \tag{2.2}$$

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. Furthermore, suppose that $\{\varepsilon_i, i \geqslant 1\}$ are independent and identically distributed random variables with mean zero and variance one. Let $\{x_i\}$ be a real sequence satisfying

$$\sum_{i=1}^{n} \{S(x_i,\theta) - S(x_i,\theta_0)\}^2 = 2nC(\lambda)|\theta-\theta_0|^{2\lambda+1}(1+o(1)) \tag{2.3}$$

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

$$nk_1|\theta_1 - \theta_2|^{2\lambda+1} \leqslant \sum_{i=1}^{n}\{S(x_i, \theta_1) - S(x_i - \theta_2)\}^2 \leqslant nk_2|\theta_1 - \theta_2|^{2\lambda+1} \tag{2.4}$$

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

$$P_{\theta_0}[n^{\rho}|\hat{\theta}_n - \theta_0| > \tau] \leqslant C\tau^{-(2\lambda+1)} \tag{2.5}$$

and

$$n^{\rho}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathscr{L}} \hat{\phi} \quad \text{as } n \to \infty, \tag{2.6}$$

where $\rho = \frac{1}{2\lambda+1}$ and $\hat{\phi}$ is the location of the minimum of the gaussian process $\{R(\phi), -\infty < \phi < \infty\}$ with

$$E[R(\phi)] = 2C(\lambda)|\phi|^{2\lambda+1} \tag{2.7}$$

and

$$Cov[R(\phi_1), R(\phi_2)] = 4C(\lambda)[|\phi_1|^{2\lambda+1} + |\phi_2|^{2\lambda+1} - |\phi_1 - \phi_2|^{2\lambda+1}]. \tag{2.8}$$

The process $\{R(\phi), -\infty < \phi < \infty\}$ can be represented in the form

$$R(\phi) = \sqrt{8C(\lambda)}\,W^H(\phi) + 2C(\lambda)|\phi|^{2\lambda+1}, \tag{2.9}$$

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

$$Cov(W^H(\phi_1), W^H(\phi_2)) = \frac{1}{2}[|\phi_1|^{2\lambda+1} + |\phi_2|^{2\lambda+1} - |\phi_1 - \phi_2|^{2\lambda+1}] \tag{2.10}$$

and $H = \lambda + \frac{1}{2}$ is the Hurst parameter.

Proof. Let $\hat{\theta}_n$ be an LSE of θ obtained by minimizing

$$Q_n(\theta) = \sum_{i=1}^{n} (Y_i - S(x_i, \theta))^2. \tag{2.11}$$

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

$$\begin{aligned} &Q_n(\theta) - Q_n(\theta_0) \\ &\quad = \sum_{i=1}^{n} (Y_i - S(x_i, \theta))^2 - \sum_{i=1}^{n} (Y_i - S(x_i, \theta_0))^2 \\ &\quad = 2\sum_{i=1}^{n} \varepsilon_i(S(x_i, \theta) - S(x_i, \theta_0)) + \sum_{i=1}^{n} (S(x_i, \theta) - S(x_i, \theta_0))^2. \end{aligned} \tag{2.12}$$

Observe that condition (2.3) implies that

$$E_{\theta_0}[Q_n(\theta) - Q_n(\theta_0)] = \sum_{i=1}^{n} (S(x_i, \theta) - S(x_i, \theta_0))^2 = 2nC(\lambda)|\theta - \theta_0|^{2\lambda+1}(1 + o(1)) \tag{2.13}$$

and

$$Var_{\theta_0}[Q_n(\theta) - Q_n(\theta_0)] = 4\sum_{i=1}^{n} (S(x_i, \theta) - S(x_i, \theta_0))^2 = 8nC(\lambda)|\theta - \theta_0|^{2\lambda+1}(1 + o(1)). \tag{2.14}$$

In general, it follows from condition (2.4) there exists $k_2 > 0$ independent of n, θ and $\theta_0 \in \Theta$ such that

$$E_{\theta_0}[Q_n(\theta) - Q_n(\theta_0)] \leqslant nk_2|\theta - \theta_0|^{2\lambda+1} \tag{2.15}$$

and

$$Var_{\theta_0}[Q_n(\theta) - Q_n(\theta_0)] \leqslant 4nk_2|\theta - \theta_0|^{2\lambda+1}. \tag{2.16}$$

Furthermore,

$$\begin{aligned} &Cov_{\theta_0}[Q_n(\theta_1) - Q_n(\theta_0), Q_n(\theta_2) - Q_n(\theta_0)] \\ &= 4\sum_{i=1}^{n} (S(x_i, \theta_1) - S(x_i, \theta_0))(S(x_i, \theta_2) - S(x_i, \theta_0)) \\ &= 4nC(\lambda)[|\theta_1 - \theta_0|^{2\lambda+1} + |\theta_2 - \theta_0|^{2\lambda+1} - |\theta_1 - \theta_2|^{2\lambda+1}](1 + o(1)) \end{aligned} \tag{2.17}$$

from the relation

$$||f||^2 + ||g||^2 - ||f - g||^2 = 2\langle f, g\rangle$$

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 [11] that there exists $\eta > 0$ such that

$$\lim_{\tau\to\infty} \limsup_{n\to\infty} P_{\theta_0}\left[\inf_{|\theta-\theta_0|>\tau n^{-\rho}} \frac{J_n(\theta)}{n|\theta - \theta_0|^{2\lambda+1}} \leqslant \eta\right] = 0, \tag{2.18}$$

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

$$P_{\theta_0}[n^{\rho}|\hat{\theta}_n - \theta_0| > \tau] \leqslant C\tau^{-(2\lambda+1)}, \tag{2.19}$$

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

$$R_n(\phi) = J_n(\theta_0 + \phi n^{-\rho}), \quad \phi \in [-\tau, \tau] \tag{2.20}$$

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

$$E[R(\phi)] = 2C(\lambda)|\phi|^{2\lambda+1} \tag{2.21}$$

and

$$Cov[R(\phi_1), R(\phi_2)] = 4C(\lambda)[|\phi_1|^{2\lambda+1} + |\phi_2|^{2\lambda+1} - |\phi_1 - \phi_2|^{2\lambda+1}]. \tag{2.22}$$

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

$$\tilde{R}(\phi) = \frac{R(\phi) - E[R(\phi)]}{\sqrt{8C(\lambda)}} \tag{2.23}$$

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

$$Cov(\tilde{R}(\phi_1), \tilde{R}(\phi_2)) = \frac{1}{2}[|\phi_1|^{2\lambda+1} + |\phi_2|^{2\lambda+1} - |\phi_1 - \phi_2|^{2\lambda+1}] \tag{2.24}$$

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

Let

$$\begin{aligned} Z_n(\phi) &= \sum_{i=1}^{n} \varepsilon_i (S(x_i, \theta_0 + \phi n^{-\rho}) - S(x_i, \theta_0)) \\ &= \sum_{i=1}^{n} a_{ni}\varepsilon_i \end{aligned} \tag{2.25}$$

and

$$T_n(\phi) = \sum_{i=1}^{n} (S(x_i, \theta_0 + \phi n^{-\rho}) - S(x_i, \theta_0))^2. \tag{2.26}$$

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

$$T_n(\phi) \to 2C(\lambda)|\phi|^{2\lambda+1} \quad \text{as } n \to \infty. \tag{2.27}$$

In addition,

$$Z_n(\phi) \xrightarrow{\mathscr{L}} N(0, 2C(\lambda)|\phi|^{2\lambda+1}) \quad \text{as } n \to \infty \tag{2.28}$$

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

$$\max_{1 \leqslant k \leqslant n} \frac{a_{nk}^2}{\sum_{i=1}^{n} a_{ni}^2} \to 0 \quad \text{as } n \to \infty. \tag{2.29}$$

This follows from a central limit theorem due to Eicker [6]. The above relation can be proved by the following arguments. Note that

$$\sum_{i=1}^{n} a_{ni}^2 = \sum_{i=1}^{n} (S(x_i, \theta_0 + \phi n^{-\rho}) - S(x_i, \theta_0))^2$$
$$= T_n(\phi) \tag{2.30}$$

which tends to $2C(\lambda)|\phi|^{2\lambda+1}$ as $n \to \infty$ by relation (2.27) and the latter in turn implies that

$$\max_{1 \leqslant i \leqslant n} (S(x_i, \theta_0 + \phi n^{-\rho}) - S(x_i, \theta_0))^2 \to 0. \tag{2.31}$$

This follows from the observation that if $\sum_{1 \leqslant k \leqslant n} a_{nk}^2 \to c > 0$ and $\max_{1 \leqslant k \leqslant N} a_{nk}^2 \to 0$ for every fixed $N \geqslant 1$, then $\max_{1 \leqslant k \leqslant n} a_{nk}^2 \to 0$ as $n \to \infty$. The above discussion proves (2.28). Similarly, it can be shown that all the finite-dimensional distributions of the process $\{Z_n(\phi), \ -\tau \leqslant \phi \leqslant \tau\}$ converge to the corresponding finite-dimensional distributions of the gaussian process $\{Z(\phi), \ -\tau \leqslant \phi \leqslant \tau\}$ with mean zero and

$$Cov[Z(\phi_1), Z(\phi_2)] = C(\lambda)[|\phi_1|^{2\lambda+1} + |\phi_2|^{2\lambda+1} - |\phi_1 - \phi_2|^{2\lambda+1}]. \tag{2.32}$$

In addition, observe that

$$E|Z_n(\phi_1) - Z_n(\phi_2)|^2 = \sum_{i=1}^{n} (S(x_i, \theta_0 + \phi_1 n^{-\rho}) - S(x_i, \theta_0 + \phi_2 n^{-\rho}))^2$$
$$\leqslant k_2|\phi_1 - \phi_2|^{2\lambda+1}, \tag{2.33}$$

where k_2 is independent of n, ϕ_1 and ϕ_2. Hence, the family of measures $\{\mu_n\}$ generated by the stochastic processes $\{Z_n(\phi), -\tau \leqslant \phi \leqslant \tau\}$ 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 $\{Z_n(\phi), \ -\tau \leqslant \phi \leqslant \tau\}$ converge weakly to the corresponding finite-dimensional distributions of the process $\{Z(\phi), \ -\tau \leqslant \phi \leqslant \tau\}$ prove that the sequence of processes $\{Z_n(\phi), \ -\tau \leqslant \phi \leqslant \tau\}$ converge weakly to the gaussian process $\{Z(\phi), \ -\tau \leqslant \phi \leqslant \tau\}$. Hence, the sequence of processes $\{R_n(\phi), \ -\tau \leqslant \phi \leqslant \tau\}$ converge weakly to the gaussian process $\{R(\phi), \ -\tau \leqslant \phi \leqslant \tau\}$ with mean function and covariance function given by (2.7) and (2.8), respectively. Applying arguments similar to those given in Prakasa Rao [11], it follows that

$$n^{\frac{1}{2\lambda+1}}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathscr{L}} \hat{\phi} \quad \text{as } n \to \infty, \tag{2.34}$$

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

$$E[R(\phi)] = 2C(\lambda)|\phi|^{2\lambda+1} \tag{2.35}$$

and

$$Cov[R(\phi_1), R(\phi_2)] = 4C(\lambda)[|\phi_1|^{2\lambda+1} + |\phi_2|^{2\lambda+1} - |\phi_1 - \phi_2|^{2\lambda+1}]. \tag{2.36}$$

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. [13]). Furthermore, for any $\tau > 0$,

$$P_{\theta_0}[n^{\rho}|\hat{\theta}_n - \theta_0| > \tau] \leqslant C\tau^{-(2\lambda+1)} \tag{3.1}$$

where the constant C is independent of n and τ.

(ii) Let

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

Conditions (2.3) and (2.4) on the sequence $\{x_i\}$ are plausible conditions that can be assumed. This can be justified by the following arguments.

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

$$f(x, \theta_0) = \begin{cases} h(x - \theta_0)\exp\{a|x - \theta_0|^{\lambda}\} & \text{if } x \leqslant \theta_0 \\ h(x - \theta_0)\exp\{b|x - \theta_0|^{\lambda}\} & \text{if } x \geqslant \theta_0. \end{cases}$$

Then 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

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

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 Lemma 4 in Dachian and Kutoyants [5]. In fact,

$$\begin{aligned} &\int_{-\infty}^{\infty} [d(x-\theta_1)|x-\theta_1|^{\lambda} - d(x-\theta_0)|x-\theta_0|^{\lambda}][d(x-\theta_2)|x-\theta_2|^{\lambda} \\ &\quad - d(x-\theta_0)|x-\theta_0|^{\lambda}]\, dx \\ &= C(\lambda)[|\theta_1|^{2\lambda+1} + |\theta_2|^{2\lambda+1} - |\theta_1-\theta_2|^{2\lambda+1}], \end{aligned} \tag{3.2}$$

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 $a = b$, is proved in Prakasa Rao [12]. For the general case, see Ibragimov and Khasminskii [8].

(iii) If $\lambda = \frac{1}{2}$ in 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

$$P_{\theta_0}(n^{1/2}|\hat{\theta}_n - \theta_0| > \tau) \leqslant C\tau^{-2} \tag{3.3}$$

and

$$n^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathscr{L}} \hat{\phi} \quad \text{as } n \to \infty, \tag{3.4}$$

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

$$R(\phi) = 2C^*\phi^2 + L\phi\psi, \tag{3.5}$$

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

$$\hat{\phi} = -\frac{L}{4C^*}\psi. \tag{3.6}$$

Combining the above remarks, we obtain that

$$\hat{\theta}_n \xrightarrow{p} \theta_0 \quad \text{as } n \to \infty, \tag{3.7}$$

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

$$P_{\theta_0}(n^{1/2}|\hat{\theta}_n - \theta_0| > \tau) \leqslant C\tau^{-2} \tag{3.8}$$

and

$$n^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathscr{L}} N(0, (2C^*)^{-1}) \quad \text{as } n \to \infty. \tag{3.9}$$

Note that the limiting distribution of the LSE $\hat{\theta}_n$ 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.

Acknowledgments

The author thanks the referee for his comments and suggestions.

References

[1] M. Akahira, K. Takeuchi, Non-regular Statistical Estimation, Springer, New York, 1995.

[2] Y. Bard, Nonlinear Parameter Estimation, Academic Press, New York, 1974.

[3] H. Bunke, W.H. Schmidt, Asymptotic results on nonlinear approximation of regression functions and weighted least squares, Math. Oper. Ser. Statist. 11 (1980) 3–32.

[4] S. Dachian, Estimation of cusp location by Poisson observations, Statist. Inference Stochastic. Process. 6 (2002) 1–14.

[5] S. Dachian, Y. Kutoyants, On cusp estimation of ergodic diffusion process, J. Statist. Plann. Inference 2002, to appear.

[6] F. Eicker, Central limit theorems for families of sequences of random variables, Ann. Math. Statist. 34 (1963) 439–466.

[7] A.R. Gallant, Nonlinear Statistical Models, Wiley, New York, 1987.

[8] I.A. Ibragimov, R.Z. Khasminskii, Statistical Estimation: Asymptotic Theory, Springer, New York, 1981.

[9] R.I. Jennrich, Asymptotic properties of nonlinear least squares estimators, Ann. Math. Statist. 40 (1969) 633–643.

[10] E. Malinvaud, The consistency of nonlinear regression, Ann. Math. Statist. 41 (1970) 956–969.

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

[12] B.L.S. Prakasa Rao, On evaluating a certain integral, Amer. Math. Monthly 75 (1968b) 55.

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

[14] B.L.S. Prakasa Rao, Asymptotic theory of least squares estimator in a nonregular nonlinear regression model, Statist. Probab. Lett. 3 (1985) 15–18.

[15] B.L.S. Prakasa Rao, Weak convergence of the least squares process in the smooth case, Statistics 17 (1986a) 505–516.

[16] B.L.S. Prakasa Rao, Weak convergence of the least squares random field in the smooth case, Statist. Decisions 4 (1986b) 363–377.

[17] B.L.S. Prakasa Rao, The weak convergence of the least squares random field and its application, Ann. Inst. Statist. Math. 38 (1986c) 475–483.

[18] B.L.S. Prakasa Rao, Asymptotic theory of estimation in nonlinear regression, in: I.B. MacNeill, G.J. Umphrey (Eds.), Advances in Statistical Sciences, Vol. I, Applied Probability, Stochastic Processes and Sampling Theory, Reidel, Dordrecht, Holland, (1987a), pp. 141–165.

[19] B.L.S. Prakasa Rao, Asymptotic Theory of Statistical Inference, Wiley, New York, 1987b.

[20] C.F. Wu, Asymptotic theory of nonlinear least squares estimation, Ann. Statist. 9 (1981) 501–513.