Azerbaijan Journal of Mathematics
V. 15, No 2, 2025, July
ISSN 2218-6816
https://doi.org/10.59849/2218-6816.2025.2.124

Gauss-Newton Method for Nonlinear Regression Model with an Increasing Number of Unknown Parameters

A.H. Hajiyev*, J. Xu

Abstract. In this paper, nonlinear regression models with an increasing number of unknown parameters are considered. The specificity of these models is that the variances of random errors are unknown and different. At each point of observation there is no more than one response which does not allow to estimate variances. Using Gauss-Newton approach, the iterative process for finding the least square estimates has been created. The conditions for convergence of the iterative process are found. It is shown that under some conditions the deviation vector of unknown parameters has Gaussian distribution, which allows to create a confidence band for unknown functions in nonlinear regression models.

Key Words and Phrases: Gauss-Newton method, nonlinear regression model, iterative process, least square estimates, asymptotic normality.

2010 Mathematics Subject Classifications: 60

1. Introduction

In statistics, nonlinear regression is a form of regression analysis where observational data are modeled by a function which is a nonlinear combination of the model parameters and depends on one or more independent variables. The data are fitted by a method of successive approximations. Nonlinear models have quite complicated structures in comparison with linear models. The examples of nonlinear regressions are the models with exponential, logarithmic, trigonometric, power functions and others. Unlike linear regression, in nonlinear regression function $f(x)$ has a nonlinear structure which does not allow to estimate unknown parameters.

*Corresponding author.

The specificity of the model considered in this paper is that variances of random errors are different and unknown, and, moreover, at each point of observation there is no more than one response which does not allow to estimate variances. Such models are typical for applications and can be used in biology, medicine, engineering and other fields. There are a lot of papers and books related to nonlinear regression models [1, 6, 11, 12], but the models with increasing number of unknown parameters and different variances are not investigated widely. In [8, 9, 13, 14] in the study of regression models with fixed number of unknown parameters, the new approaches for deriving the least square estimators (l.s.e.) and the estimators of the elements of covariance matrix of a deviation vector of unknown parameters are suggested. Such approaches allow to create a confidence band for unknown functions in regression models. It is necessary to underline two papers [3,14], where the new approaches for investigation of regression models with different and unknown values of the variances of random errors are suggested.

One of the effective approaches for investigation of nonlinear regression models is Gauss-Newton method [5] which allows to create an iterative process for calculation of least square estimators. Nevertheless, for regression models with an increasing number of unknown parameters it is necessary to establish the needed conditions for a convergence of the iterative process created by Gauss – Newton method. Some approaches for investigation of nonlinear regression models with missing data are suggested in [16,17].

2. The properties of least square estimates (l.s.e.)

Consider the following regression model:

$$y_i = \eta\left(x_i,\ \theta^*\right) + \varepsilon_i \qquad i = \overline{1, N}, \tag{1}$$

where $\eta\left(x_i,\ \theta^*\right)$ is a function nonlinear in θ, and θ^* is a true value of the parameter θ.

We assume that $\eta\left(x,\ \theta\right);\ \frac{\partial\eta(x,\ \theta)}{\partial\theta};\ \frac{\partial^2\eta(x,\ \theta)}{\partial\theta_i\partial\theta_j},\ i,j = \overline{1,m}$, are the bounded functions continues in $(x,\ \theta)$ and $\theta \in \Theta$ is a compact set,

$$E\varepsilon_i = 0, E\varepsilon_i\varepsilon_j = r_{ij}, r_{ii} = \sigma_i^2, 0 < \sigma_*^2 \leq \sigma_i^2 \leq \sigma_0^2. \tag{2}$$

Denote

$$f_{ij}\left(\theta\right) = \frac{\partial\eta\left(x_j,\ \theta\right)}{\partial\theta_i},\ i = \overline{1,m}; j = \overline{1,N};$$

$\widehat{\theta}$ means l.s.e.

$F_N(\theta)$ is the matrix with the elements $f_{ij}\left(\theta\right);$

$0<\lambda_1^{(N)}(\theta)\le\lambda_2^{(N)}(\theta)\le\cdots\le\lambda_m^{(N)}(\theta)$ are the eigen values of the matrix $\left[\frac{F_N^T(\theta)F_N(\theta)}{N}\right]$, for which the following relations hold:

$$0<\lambda_1^{(N)}(\theta)\le\lambda_2^{(N)}(\theta)\le\cdots\le\lambda_m^{(N)}(\theta).$$

$B(r)$ is the sphere with the radius $r>0$ centered at the point θ^*.

To find l.s.e., we can use the iterative process [4] according to the Gauss-Newton method:

$$\theta(s+1)=\theta(s)+\left[F_N^T(\theta(s))\cdot F_N(\theta(s))\right]^{-1}F_N^T(\theta(s))\left(y-\eta\left(x,\ \theta(s)\right)\right), \quad (3)$$

All values in (3) are defined above. The main problem is the convergence of iterative process (3). Throughout this paper we assume that $\theta\in B(r)$.

Let us rewrite the expression (3) as follows:

$$\theta_N(s+1)=U(\theta_N(s))=\theta_N(s)+A_N(\theta_N(s))\cdot\delta_N(\theta_N(s)), \quad (4)$$

where

$$A_N(\theta_N(s))=\left[\frac{F_N^T(\theta(s))\cdot F_N(\theta(s))}{N}\right]^{-1}\frac{F_N^T(\theta(s))}{N} \quad (5)$$

$$\delta_N(\theta(s))=y-\eta(x,\ \theta(s));\quad \delta^*=y-\eta(x,\theta^*)=\varepsilon.$$

Let's state some preliminary auxiliary results.

$$A_{n_1}=\left(a^{(n_1)}\right)_{ij};\ B_{n_2}=\left(b^{(n_2)}\right)_{ij}:i=\overline{1,m};j=\overline{1,N};\quad C=(c)_{ij}:i,j=\overline{1,m}.$$

Denote

$$D=A_1A_2^T\ldots A_{k-1}^TA_k\cdot C^nB_1^TB_2\ldots B_{l-1}B_l^T,$$

where $k=2k_1+1; l=2l_1+1; n,\ l_1,\ k_1$ are integer numbers.

Lemma 1. *If there exists some constant C_1 such that*

$$\left|a_{ij}^{(n_1)}\right|\le C_1,\ \left|b_{ij}^{(n_2)}\right|\le C_1,\ \max_{1\le i\le m}\sum_{j=}^{m}|c_{ij}|\le C_1,$$

then the following inequality holds:

$$E\left(\varepsilon^TD\varepsilon\right)\le C_1^nm^{k_1+l_1+1}N^{k_1+l_1-1}\left(C_1N+\sum_{i\ne j}|r_{ij}|\right). \quad (6)$$

Proof. Denote by d_{ij} $i,j=\overline{1,N}$, the elements of the matrix D. Then

$$E\left(\varepsilon^T D\varepsilon\right)=E\sum_{i=1}^{N}\sum_{j=1}^{N}d_{ij}\varepsilon_i\varepsilon_j=\sum_{i=1}^{N}d_{ii}E\varepsilon_i^2+\sum_{i\neq j}d_{ij}\varepsilon_i\varepsilon_j\leq\leq\sigma_o^2 trD+$$

$$+\sum_{i\neq j}d_{ij}\left|r_{ij}\right|. \tag{7}$$

Using the conditions of Lemma 1, we have

$$d_{ij}=\sum_{i_1=1}^{N}\cdots\sum_{i_{k_1}=1}^{N}\sum_{j_1=1}^{N}\cdots\sum_{i_{k_1+1}=1}^{m}\sum_{p_1=1}^{m}\cdots\sum_{p_n=1}^{m}\sum_{r_1=1}^{N}\cdots\sum_{r_{l_1}=1}^{N}\sum_{s_1=1}^{m}\cdots$$

$$\cdots\sum_{s_{l_1}=1}^{m}a_{ij_1}^{(1)}a_{j_1i_1}^{(2)^T}\ldots a_{i_{k_1}p_1}^{(k_1)}c_{p_1p_2}\ldots c_{p_{n-1}n}\times$$

$$\times b_{p_nr_nj_1}^{(1)^T}b_{r_1s_1}^{(2)}\ldots b_{s_{l_1}j}^{(l)}\leq C_1^nN^{k_1}m^{k_1+1}N^{l_1}m^{l_1+1}=C_1N^{k_1+l_1}m^{k_1+l_1+2}$$

$$trD=\sum_{i=1}^{m}d_{ii}\leq C_1^{n+1}N^{k_1+l_1}m^{k_1+l_1+2}N.$$

Then from here and (7) it follows (6), i.e. the assertion of Lemma 1. ◀

Denote

$$\zeta_{N,r}^n\left(\theta\right)=m\cdot\frac{\partial A_N\left(\theta\right)}{\partial\theta_n}\cdot\varepsilon,\ \theta\in B\left(r\right),\ n=1,2,\ldots,m.$$

Theorem 1. *Assume that the conditions (1)-(3) hold. If* $\exists$ N *such that*

$$\frac{m^4\sqrt{m}}{N(\lambda_1\left(\theta\right))^3}+\frac{m^4}{N^2(\lambda_1\left(\theta\right))^3}\sum_{i\neq j}\left|r_{ij}\right|\to 0\quad as\ r\to 0 \tag{8}$$

and

$$\frac{m^5}{N(\lambda_1\left(\theta\right))^4}+\frac{m^5}{N^2(\lambda_1\left(\theta\right))^4}\sum_{i\neq j}\left|r_{ij}\right|\to 0\quad as\ r\to 0, \tag{9}$$

then

$$\zeta_{N,r}^{(n)}\xrightarrow{P}0\quad as\ r\to 0. \tag{10}$$

Proof. From the matrix analysis of [2] we have

$$\frac{\partial A_N(\theta)}{\partial \theta_n}\varepsilon = \frac{1}{N}\left[\frac{F_N^T(\theta(s))\,F_N(\theta(s))}{N}\right]^{-1}\frac{\partial F_N^T(\theta)}{\partial \theta_n} + \frac{1}{N}\left[\frac{F_N^T(\theta)\,F_N(\theta)}{N}\right]^{-2}\times$$

$$\times\left[\left(\frac{\partial F_N^T(\theta)}{\partial \theta_n}F_N(\theta) + F_N^T(\theta)\frac{\partial F_N(\theta)}{\partial \theta_n}\right)/N\right]\cdot\varepsilon. \tag{11}$$

Denote by $I_k,\ i = 1,\ 2,\ 3,$ the k-th term in (10). Then

$$I_1 \le m\cdot\frac{C}{N^2}\cdot m\left(\frac{\sqrt{m}}{\lambda_1(\theta)}\right)^2\cdot\left(N + \sum_{i\ne j}|r_{ij}|\right) =$$

$$= C_1\left(\frac{m^3}{N(\lambda_1(\theta))^2} + \frac{m^3}{N^2(\lambda_1(\theta))^2}\sum_{i\ne j}|r_{ij}|\right) \to 0 \quad as \quad r \to 0$$

according to the condition (9).

Opening parenthesis in I_2 and denoting by $J_{2,i}(\theta)$ the i-th term in I_2, we have

$$J_{2,1}(\theta) = \frac{m}{N^2}E\left[\varepsilon^T\frac{\partial F_N(\theta)}{\partial \theta_n}\left(\frac{F_N^T(\theta)\,F_N(\theta)}{N}\right)^{-3}\frac{\partial F_N(\theta)}{\partial \theta_n}F_N(\theta)\,F_N^T(\theta)\,\varepsilon\right].$$

Similarly, we have

$$J_{2,1}(\theta) \le \frac{C}{N^2}\cdot m^3\left(\frac{\sqrt{m}}{\lambda_1(\theta)}\right)^3\left(N + \sum_{i\ne j}|r_{ij}|\right) =$$

$$= C\cdot\frac{m^4\sqrt{m}}{N(\lambda_1(\theta))^3} + \frac{m^4\sqrt{m}}{(\lambda_1(\theta))^3}\cdot\frac{1}{N^2}\sum_{i\ne j}|r_{ij}| \to 0, r \to 0,$$

according to the condition (8).

By the same way we can prove $J_{2,2}(\theta) \to 0$ and $I_3 \to 0$ as $r \to 0$.

Let us open parenthesis in I_4. Denote by $J_{4,i}$ the i-th term in I_4:

$$J_{4,1} = \frac{m}{N^2}E\left(\varepsilon^T F_N(\theta)\frac{\partial F_N^T(\theta)}{\partial \theta_n}F_N(\theta)\times\right.$$

$$\left.\times\left(\frac{F_N^T(\theta)\,F_N(\theta)}{N}\right)^{-4}\frac{\partial F_N^T(\theta)}{\partial \theta_n}F_N(\theta)\,F_N^T(\theta)\,\varepsilon\right).$$

According to Lemma 1 and the conditions of Theorem 1, we have

$$J_{4,1} = \frac{1}{N^2} \cdot C \cdot m^3 \left(\frac{\sqrt{m}}{\lambda_1(\theta)} \right)^4 \left(N + \sum_{i \neq j} |r_{ij}| \right) =$$

$$= \frac{C}{N} \cdot \frac{m^5}{\lambda_1^4(\theta)} + \frac{m^5}{N^2 \lambda_1^4(\theta)} \sum_{i \neq j} |r_{ij}| \to 0, r \to 0.$$

Other terms $J_{4,i}$, $i = 2, 3, 4$, have the same form as $J_{4,i}$. Hence, $J_4(\theta) \to 0$ as $r \to 0$. Then from Chebyshev's inequality we have

$$P\left\{ \left\| m \cdot \frac{\partial A_N(\theta)}{\partial \theta_n} \varepsilon \right\|^2 > a \right\} \le \frac{E\left(\varepsilon^T \frac{\partial A_N^T(\theta)}{\partial \theta_n} \cdot \frac{\partial A_N(\theta)}{\partial \theta_n} \cdot \varepsilon \right)}{a^2} \le$$

$$\le \frac{I_1 + I_2 + I_3 + I_4}{a^2} \to 0, \; r \to 0,$$

which proves Theorem 1. ◄

Remark 1. *The conditions* (8) *and* (9) *can be rewritten in a different form. If* $m/\lambda_1(\theta) \to 0$, *and* $\max_{i,j} |r_{ij}| \le C < \infty$ *as* $r \to 0$, *then* (10), *i.e. the statement of Theorem 1 stays true.*

Denote

$$L_n = \frac{\partial U_N(\theta)}{\partial \theta_n}; \; \tau_N(r) = \max_{p=1,\dots,m} \sup \|L_p\| \; .$$

Theorem 2. *Under the conditions of Theorem 1, the following relation is true:*

$$m \cdot \tau_N(r) \xrightarrow{P} 0 \text{ as } r \to 0.$$

Proof. Let $\overline{f}_{ij}(\theta)$ be the elements of the matrix $\left(\frac{F_N^T(\theta) F_N(\theta)}{N} \right)^{-4}$, $i, j = \overline{1, m}$; and $\partial f_{kl}(\theta)$ be the elements of the matrix $\left(\frac{\partial F_N(\theta)}{\partial \theta_n} \right)$, $k = \overline{1, m}; l = \overline{1, N}$.

Then from expression (5) it follows

$$E\left\| m \cdot \frac{\partial A_N(\theta)}{\partial \theta_n} \cdot \varepsilon \right\|^2 = m \cdot E\left(\varepsilon^T \frac{\partial A_N^T(\theta)}{\partial \theta_n} \cdot \frac{\partial A_N(\theta)}{\partial \theta_n} \cdot \varepsilon \right) =$$

$$\frac{m^2}{N} E\left[\varepsilon^T \frac{\partial F_N(\theta)}{\partial \theta_n} \left(\frac{F_N^T(\theta) F_N(\theta)}{N} \right)^{-1} \left(\frac{F_N^T(\theta) F_N(\theta)}{N} \right)^{-1} \frac{\partial F_N(\theta)}{\partial \theta_n} \varepsilon \right] +$$

$$\frac{m}{N^2} E\left[\varepsilon^T \frac{\partial F_N(\theta)}{\partial \theta_n} \left(\frac{F_N^T(\theta) F_N(\theta)}{N} \right)^{-3} \left(\frac{\partial F_N(\theta)}{\partial \theta_n} \cdot F_N(\theta) + F_N^T(\theta) \frac{\partial F_N(\theta)}{\partial \theta_n} \right) F_N^T(\theta) \cdot \varepsilon \right] +$$

$$\frac{m^2}{N}E\left[\varepsilon^T F_N(\theta)\left(\frac{\partial F_N^T(\theta)}{\partial\theta_n}\cdot F_N(\theta)+F_N^T(\theta)\frac{\partial F_N(\theta)}{\partial\theta_n}\right)\left(\frac{F_N^T(\theta)F_N(\theta)}{N}\right)^{-2}\left(\frac{F_N^T(\theta)F_N(\theta)}{N}\right)^{-2}\right.$$

$$\left.\left(\frac{\partial F_N^T(\theta)}{\partial\theta_n}\cdot F_N(\theta)+F_N(\theta)\frac{\partial F_N(\theta)}{\partial\theta_n}\right)\right]F_N^T(\theta)\cdot\varepsilon. \tag{12}$$

Denote by I_4, $i=\overline{1,4}$, the i-th term in (12). Using the condition (5) and the statement of Lemma 1, for elements of the matrix $\frac{F_N^T(\theta)F_N(\theta)}{N}$ we have

$$\frac{\partial\delta_N(\theta)}{\partial\theta_n}=-\frac{\partial\eta(x,\theta)}{\partial\theta_n}=-F_N(\theta).$$

Then from (10) it follows

$$m\cdot\frac{\partial U_N(\theta)}{\partial\theta_n}=\left[I_n+\frac{\partial A_N(\theta)}{\partial\theta_n}\delta_N(\theta)+A_N(\theta)\cdot\frac{\partial\delta_N(\theta)}{\partial\theta_n}\right]m=$$

$$=m\cdot\left[\frac{\partial A_N(\theta)}{\partial\theta_n}\delta(\theta^*)+\frac{\partial A_N(\theta)}{\partial\theta_n}(\delta_N(\theta)-\delta(\theta^*)\right]=$$

$$=\frac{\partial A_N(\theta)}{\partial\theta_n}\cdot\varepsilon\cdot m+\frac{\partial A_N(\theta)}{\partial\theta_n}\cdot\Delta\eta(x,\theta,\theta^*), \tag{13}$$

where $\Delta\eta(x,\theta,\theta^*)=\eta(x,\theta^*)-\eta(x,\theta)$.

According to Theorem 1, the first term in (13) goes to zero as $r\to 0$.

The second term goes to zero because $\Delta\eta(x,\theta,\theta^*)\to 0$ as $r\to 0$ uniformly in x. Hence, Theorem 2 is proved. ◄

Denote

$$\rho_N(\theta)=U_N(\theta)-\theta,\ \rho^*=\rho(\theta^*).$$

Theorem 3. *Assume*

$$\frac{m}{\lambda_1(\theta)}+\frac{m}{N(\lambda_1(\theta))^2}\sum_{i\neq j}|r_{ij}|\le A,$$

where A is some constant. Then

$$\lim_{N\to\infty}\lim_{k\to\infty}P\left\{\sqrt{N}\cdot\rho^*>k\right\}=0.$$

Proof. Consider

$$NE\|\rho_N(\theta)\|^2=E(U_N(\theta)-\theta)^T(U_N(\theta)-\theta)=NE\varepsilon^T A_N^T(\theta)A_N(\theta)\varepsilon=$$

$$=E\varepsilon^T F_N(\theta)\cdot\left(\frac{F_N^T(\theta)F_N(\theta)}{N}\right)^{-1}\left(\frac{F_N^T(\theta)F_N(\theta)}{N}\right)^{-1}\frac{F_N^T(\theta)}{N}\varepsilon. \tag{14}$$

As the elements of the matrix $F_N(\theta)$ are bounded, using the expression (9) and Lemma 1, according to the condition of Theorem 3, we get

$$NE\|\rho\|^2 \le \sigma_0^2 tr\left(\frac{F_N^T(\theta)\,F_N(\theta)}{N}\right)^{-1} + \frac{m}{N(\lambda_1(r))^2}\sum_{i\neq j}|r_{ij}| =$$

$$= \frac{c_1\cdot m}{\lambda_1(\theta)} + \frac{m}{N(\lambda_1(\theta))^2}\sum_{i\neq j}|r_{ij}| < A,\ \forall\theta\in B(r).$$

As $\theta^*\in B(r)$, we have $NE\|\rho^*\|^2 < A$.

Using Chebyshev's inequality, we get the statement of Theorem 3. ◄

Let us pass on to the convergence of the iterative process (4).

Theorem 4. *Assume* $\theta(0)\in B(r)$ *and*

$$\tau_N(r) + \frac{\|\rho^*\|}{r} < 1.$$

Then, by Theorem 3, there exists such a random variable $\widehat{\theta}_N$ that $\left(\widehat{\theta}_N - \widehat{\theta}_N(s)\right)\xrightarrow{P}0$, as $s\to\infty$.

Proof. As the partial derivative $\frac{\partial U_N(\theta)}{\partial\theta}$ exists then using Lagrange's formula we have

$$\|U_N(\theta) - U_N(\theta^*)\| \le \left\|\frac{\partial U_N\left(\widetilde{\theta}\right)}{\partial\theta}\right\|\cdot\|\theta-\theta^*\|,\ \theta\in B(r),\ \widetilde{\theta}\in B(r)$$

Using Theorem 2, for $\forall\widetilde{\theta}\in B(r)$ we have

$$\left\|\frac{\partial U_N\left(\widetilde{\theta}\right)}{\partial\theta}\right\| = \left\|\sum_{p=1}^{m}\frac{\partial U_N\left(\widetilde{\theta}\right)}{\partial\theta_p}\right\| \le m\cdot\sup_{1\le p\le m}\left\|\frac{\partial U_N\left(\widetilde{\theta}\right)}{\partial\theta_p}\right\| \to 0$$

as $r\to 0$.

Hence, if $\theta(0)\in B(r)$, then there exists N such that starting from some r the mapping $U_N(\theta)$ is compressive. On the other hand,

$$\|U_N(\theta)-\theta^*\| \le \|U_N(\theta)-U_N(\theta^*)\| + \|U_N(\theta^*)-\theta^*\| \le \tau_N(r)\cdot r + \|\rho^*\| \le r,$$

$$\theta\in B(r)$$

according to Theorem 2. Thus, the mapping $U_N(\theta)$ transfers the sphere $B(r)$ into itself. As $B(r)$ is a complete metric space, according to the principle of compression mapping [9] for all $\theta\ (0) \in B(r)$, $r \to 0$, $s \to \infty$ the equation $U_N(\theta) = \theta$ has a unique solution, denoted by $\widehat{\theta}$. As $U_N(\theta(s-1)) = \theta\ (s)$ and $\left\| U_N(\theta(s-1)) - u(\widehat{\theta}) \right\| < r \to 0$ as $s \to \infty$, it follows that $\left\| \theta_N(s) - \widehat{\theta}_N \right\| \xrightarrow{P} 0$ in probability as $s \to \infty$, $r \to 0$. ◄

Remark 2. *Theorem 4 has the following explanation. The l.s.e. is a solution of the iterative process, which depends on the iteration. If from some practical point of view, we can choose $\theta_0 \in B(r)$, then for $s \in \infty$ (where s is a number of iterative steps), the iterative process converges to some θ_N and for $N \to \infty$ it tends to some $\theta(\infty)$. As l.s.e. i.e. as a true value of unknown parameters, we can take $\theta(\infty)$*

Theorem 5. *If $\tau_N(r) + \frac{\|\rho^*\|}{r} < 1$, then under the conditions of Theorem 4 the following expression holds:*

$$\sup_{\theta(0)\in B(r)} \left\| \theta_N(s) - \hat{\theta}_N \right\| \le r \frac{(\tau_N(r))^s}{1-\tau_N(r)}.$$

Proof. Consider $s < t$. Then, according to the inequality of the principle of compression mapping [10], we have

$$\sup_{\theta(0)\in B(r)} \|\theta(s) - \theta(t)\| \le \frac{(\tau_N(r))^s}{1-\tau_N(r)} \|\theta(0) - \theta(1)\|. \tag{15}$$

As the mapping $U_N(\theta)$ also transfers the sphere $B(r)$ into itself, from (15) it follows

$$\sup_{\theta(0)\in B(r)} \|\theta(s) - \theta(t)\| \le \frac{(\tau_N(r))^s}{1-\tau_N(r)} \cdot r.$$

As $\theta(t) \in B(r)$, we apply the operator $U_N(\theta)$, to $\theta(t)$ and according to Theorem 4 we get the limit point $\widehat{\theta}$, i.e.

$$\sup_{\theta(0)\in B(r)} \left\| \theta_N(s) - \hat{\theta}_N \right\| \le \frac{(\tau_N(r))^s}{1-\tau_N(r)} \cdot r.$$

◄

Theorem 6. *Under the conditions of Theorem 5, the random variable*

$$\sqrt{N}\left(\widehat{\theta}_N - \theta^*\right)$$

is bounded in probability as $N \to \infty$, i.e. the estimator $\widehat{\theta}$ is $\sqrt{N}$-consistent.

Proof. Consider $r = \frac{2k}{\sqrt{N}}$. According to Theorem 2, we have $\tau(r) \to 0,\ N \to \infty$ in probability, and according to Theorem 5, the random variable $\sqrt{N} \cdot \rho^*$ is bounded in probability. Then, choosing

$$\tau(r) = \frac{1}{2}; r = \frac{2k}{\sqrt{N}};\ \|\rho^*\| < \frac{k}{\sqrt{N}},$$

we get

$$\tau_N(r) + \frac{\|\rho^*\|}{r} < \frac{1}{2} + \frac{\|\rho^*\| \cdot \sqrt{N}}{2k} < \frac{1}{2}.$$

Taking into consideration

$$\left\|\widehat{\theta}_N - \theta^*\right\| \le \sup_{\theta(0)\in B(r)} \left\|\widehat{\theta} - \theta(s)\right\|$$

and using the conditions of Theorem 5, we have

$$\lim_{k\to\infty} \lim_{N\to\infty} P\left\{\left\|\widehat{\theta}_N - \theta^*\right\| > \frac{k}{\sqrt{N}}\right\} = 0\ .$$

Theorem 7. *Let ε_i, $i = 1, 2, \ldots$; be the sequence of independent random variables. If*

$$\frac{m^3}{N \cdot \lambda_1^2(\theta)} \to 0$$

as $N \to \infty$, $\forall \theta \in B(r)$, then $\sqrt{N} \cdot \rho^* \to N\{0, \Sigma(\theta^*)\}$ as $N \to \infty$, where

$$\Sigma(\theta^*) = \left(\frac{F_N^T(\theta^*) F_N(\theta^*)}{N}\right)^{-1} \frac{F_N^T(\theta^*) I(\sigma^2) F_N(\theta^*)}{N} \left(\frac{F_N^T(\theta^*) F_N(\theta^*)}{N}\right)^{-1}, \quad (16)$$

$$I(\theta^2) = E\varepsilon\varepsilon^T = \begin{pmatrix} \sigma & 0, & .... & ,0 \\ 0 & \sigma_2^2 0, & ... & ,0 \\ 0 & 0 & \sigma_i^2 & ,0 \\ 0 & 0, & .... & ,\sigma_N^2 \end{pmatrix}.$$

Proof. Introduce the nonzero vector of the size m :

$$l = (l_1, l_2, \ldots, l_m)^T.$$

$$\sqrt{N} l^T \rho^* = \sqrt{N} l^T \left(\frac{F_N^T(\theta^*) F_N(\theta^*)}{N}\right)^{-1} \frac{F_N^T(\theta^*)}{N} (y - \eta(x,\ \theta^*)).$$

Then

$$\sqrt{N} l^T \rho^* = \frac{1}{N} \sum_{k=1}^{N} \sum_{i=1}^{m} \sum_{j=1}^{m} \sqrt{N} l_i \overline{f}_{ij} f_{jk} \left(\theta^*\right) \varepsilon_k,$$

where $f_{jk}(\theta)$ are the elements of the matrix $F_N^T(\theta^*)$, $j = \overline{1, N}; k = \overline{1, m}$; and x_{ij} are the elements of the matrix $\left(\frac{F_N^T(\theta^*) F_N(\theta^*)}{N}\right)^{-1}$, $i, j = \overline{1, m}$;

Then $E\sqrt{N} l^T \rho^* = 0$ according to Theorem 3. Consider

$$E\left(\sqrt{N} l^T \rho^*\right)^2 = \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{i=1}^{m} \sum_{j=1}^{m} l_i x_{ij} f_{jk} \right)^2 \sigma_i^2 \leq$$

$$\leq \frac{\sigma_0^2}{N} \left(\frac{\sqrt{m}}{\lambda_1(\theta)} \right)^2 C_1^2 \cdot N \cdot \left(\sum_{i=1}^{m} l_i \right)^2 \leq C_1^2 \cdot \sigma_0^2 \cdot \frac{m^3}{\lambda_1^2(\theta)} \leq C_2 \cdot \frac{m^3}{\lambda_1^2(\theta)}, \ \forall \theta \in B(r),$$

which is bounded according to Theorem 6.

According to the central limit theorem [4], the random variable $\sqrt{N} l^T \rho^*$ has asymptotically Gaussian distribution and

$$E\left\|\sqrt{N} l^T \rho^*\right\| = N \cdot E l^T \rho^* \rho^{*T} l = l^T \left(\frac{F_N^T(\theta^*) F_N(\theta^*)}{N} \right)^{-1} \times$$

$$\times \frac{F_N^T(\theta^*) I(\sigma^2) F_N(\theta^*)}{N} \left(\frac{F_N^T(\theta^*) F_N(\theta^*)}{N} \right)^{-1} \cdot l = l^T \Sigma(\theta^*) \cdot l.$$

According to the theorem on criterion of normality [15], it follows

$$\sqrt{N} \rho^* \to N\left(0, \Sigma(\theta^*)\right).$$

Let us find a minimal number of steps when asymptotic normality is held. Assume

$$\tau_N(r) + \frac{\|\rho^*\|}{r} < a < 1, \tag{17}$$

$$\frac{r \cdot a^b}{1-a} < \frac{k}{\sqrt{N}}. \tag{18}$$

Such k can be found always because the random variable $\sqrt{N} \left\|\widehat{\theta}_N - \theta^*\right\|$ is bounded according to Theorem 6. ◀

Lemma 2. *If* (17) *is true and*

$$s \geq \frac{ln\left[(1-a)\cdot k\right] - ln\left(r\cdot\sqrt{N}\right)}{lna}, \tag{19}$$

then under the conditions of Theorem 6 the following relation holds:

$$\left(\theta\left(s\right)-\theta^{*}\right)\xrightarrow{P}N(0,\ \cdot)\quad as\ N\rightarrow\infty.$$

Proof. The statement of Lemma 2 follows from the statement of Theorem 6 and from the relations (18) and (19).

Statement 1. *The random variable $\widehat{\theta}$ satisfies the following equation:*

$$A_N\left(\widehat{\theta}\right)\cdot\delta_N\left(\widehat{\theta}\right)=0,$$

which is called a normal equation for the l.s.e.

As $\theta\left(s+1\right)=u_N\left(\theta\left(s\right)\right)=\theta\left(s\right)+A_N\left(\theta\left(s\right)\right)\cdot\delta_N\left(\theta\left(s\right)\right)$, for $s\rightarrow\infty$ we have

$$\widehat{\theta}=\widehat{\theta}+A_N\left(\widehat{\theta}\right)\cdot\delta_N\left(\widehat{\theta}\right).$$

Then it follows $A_N\left(\widehat{\theta}\right)\cdot\delta_N\left(\widehat{\theta}\right)=0.$

Theorem 8. *If the random variable $\sqrt{N}\left(\widehat{\theta}-\theta^{*}\right)$ is bounded in probability and the conditions of Lemmas 1 and 2 and Theorems 1-3 hold, then*

$$\sqrt{N}\left(\widehat{\theta}-\theta^{*}\right)\rightarrow N\left(0,\Sigma(\theta^{*})\right)$$

Proof. According to Statement 1, we have

$$A_N\left(\widehat{\theta}\right)\cdot\delta_N\left(\widehat{\theta}\right)=0. \tag{20}$$

Then according to (20) we have

$$A_N\left(\widehat{\theta}\right)\cdot\delta_N\left(\widehat{\theta}\right)=$$

$$\left[A_N\left(\theta^{*}\right)+A_N\left(\widehat{\theta}\right)-A_N\left(\theta^{*}\right)\right]\left[y-\eta\left(x,\theta^{*}\right)+\eta\left(x,\theta^{*}\right)-\eta\left(x,\widehat{\theta}\right)\right]=$$

$$A_N\left(\theta^{*}\right)\left(y-\eta\left(x,\theta^{*}\right)\right)+A_N\left(\theta^{*}\right)\left(\eta\left(x,\theta^{*}\right)-\eta\left(x,\widehat{\theta}\right)\right)+$$

$$\left[A_N\left(\widehat{\theta}\right)-A_N\left(\theta^*\right)\right]\left(y-\eta\left(x,\theta^*\right)\right)+\left[A_N\left(\widehat{\theta}\right)-A_N\left(\theta^*\right)\right]\left(\eta\left(x,\theta^*\right)-\eta\left(x,\widehat{\theta}\right)\right)=$$

$$\rho^*+\left[A_N\left(\widehat{\theta}\right)-A_N\left(\theta^*\right)\right]\cdot\delta^*+A_N\left(\widehat{\theta}\right)\cdot\Delta\eta\left(x,\theta^*,\widehat{\theta}\right)=0. \tag{21}$$

For $\left[A_N\left(\widehat{\theta}\right)-A_N\left(\theta^*\right)\right]$ and $\Delta\eta\left(x,\theta^*,\theta(\infty)\right)$ the following identities hold:

$$A_N\left(\widehat{\theta}\right)-A_N\left(\theta^*\right)=$$

$$\left[\frac{\partial A_N\left(\theta^*\right)}{\partial\theta}+\int_0^1\left(\frac{\partial A_N\left(\theta^*+z\left(\widehat{\theta}-\theta^*\right)\right)}{\partial\theta}-\frac{\partial A_N\left(\theta^*\right)}{\partial\theta}\right)dz\cdot\left(\widehat{\theta}-\theta^*\right)\right] \tag{22}$$

$$\eta\left(\widehat{\theta}\right)-\eta\left(\theta^*\right)=\left[\frac{\partial\eta\left(\theta^*\right)}{\partial\theta}+\int_0^1\left(\frac{\partial\eta\left(\theta^*+z\left(\widehat{\theta}-\theta^*\right)\right)}{\partial\theta}-\frac{\partial\eta\left(\theta^*\right)}{\partial\theta}\right)dz\right]\left(\widehat{\theta}-\theta^*\right) \tag{23}$$

Substituting (23) in (22) we get

$$\rho^*+\frac{\partial A_N\left(\theta^*\right)}{\partial\theta}\delta_N^*\left(\widehat{\theta}-\theta^*\right)\int_0^1\frac{\partial A_N\left(\theta^*+z\left(\widehat{\theta}-\theta^*\right)\right)}{\partial\theta}\delta_N^*\left(\widehat{\theta}-\theta^*\right)-$$

$$-\frac{\partial A_N\left(\theta^*\right)}{\partial\theta}\delta_N^*\left(\widehat{\theta}-\theta^*\right)+A_N\left(\widehat{\theta}\right)\frac{\partial\eta\left(\theta^*\right)}{\partial\theta}\left(\widehat{\theta}-\theta^*\right)+$$

$$+A_N\left(\widehat{\theta}\right)\int_0^1\frac{\partial\eta}{\partial\theta}\left(\theta^*+z\left(\widehat{\theta}-\theta^*\right)\right)\left(\widehat{\theta}-\theta^*\right)-$$

$$-A_N\left(\widehat{\theta}\right)\frac{\partial\eta\left(\widehat{\theta}\right)}{\partial\theta}\left(\widehat{\theta}-\theta^*\right)-A_N\left(\widehat{\theta}\right)\frac{\partial\eta\left(\theta^*\right)}{\partial\theta}\left(\widehat{\theta}-\theta^*\right)=0.$$

As

$$A_N\left(\widehat{\theta}\right)\cdot\frac{\partial\eta\left(\widehat{\theta}\right)}{\partial\theta}=I_m,$$

we have $\sqrt{N}\left(\widehat{\theta}-\theta^*\right)=\sqrt{N}\cdot\rho^*+\sqrt{N}\cdot M_1+\sqrt{N}\cdot M_2$, where

$$M_1=\int_0^1\frac{\partial A_N\left(\theta^*+z\left(\widehat{\theta}-\theta^*\right)\right)}{\partial\theta}\cdot\delta_N^*dz\left(\widehat{\theta}-\theta^*\right),$$

$$M_2=A_N\left(\widehat{\theta}\right)\int_0^1\frac{\partial\eta\left(\theta^*+z\left(\widehat{\theta}-\theta^*\right)\right)}{\partial\theta}dz\left(\widehat{\theta}-\theta^*\right).$$

As the random variable $\sqrt{N}\left(\widehat{\theta}-\theta^{*}\right)$ is bounded in probability, according to Theorem 1 we have

$$\frac{\partial A_N(\theta)}{\partial \theta} \cdot \delta_N^{*} \xrightarrow{P} 0, \qquad N \to \infty.$$

Then we get $\sqrt{N} \cdot M_1 \xrightarrow{P} 0$ as $N \to \infty$. Similarly we can prove that

$$\sqrt{N} \cdot M_2 \xrightarrow{P} 0 \ \ as \ N \to \infty.$$

Now using Theorem 7 we get $\sqrt{N}\left(\widehat{\theta}-\theta^{*}\right) \to N\left[0,\ \Sigma(\theta^{*})\right]$. ◄

The next step is an estimation of the elements of covariance matrix of the deviation vector $\sqrt{N}(\widehat{\theta}-\theta)$, where $\widehat{\theta}$ is the l.s.e., θ^{*} is a true value of unknown parameter, and N is a number of observations. For estimation of the elements of covariance matrix

$$C_N = NE(\widehat{\theta}-\theta)(\widehat{\theta}-\theta))^{T},$$

we can use the approaches suggested in [7, 8, 14], but detailed research will be the subject of our next publication.

3. Conclusion

For regression models with an increasing number of unknown parameters and different variances of random errors, the new method for calculation of the least square estimators is suggested. The specificity of these models is that at each point of observation there is no more than one response which causes difficulties with estimation of variance of a random error. Our method is based on the usage of Gauss-Newton approach and shows that under some conditions the deviation vector has a Gaussian distribution. This fact allows to create a confidence ellipsoid and further a confidence band for unknown function in nonlinear regression models. Such models are typical for applications and allow to solve problems in biology, medicine, engineering, psychology and other fields.

References

[1] D.G. Chen, J.K. Chen, *Nonlinear Regression Modeling. In: Statistical Regression Modeling with R. Emerging Topics in Statistics and Biostatistics*, Springer, Cham., 2021. https://doi.org/10.1007/978-3-030-67583-7_6.

[2] R.A. Bellman, *Introduction to matrix analysis*, NY, McGraw-Hill, 1962.

[3] Yu.K. Belyaev, A.H. Hajiyev, *Statistical data analysis*, Izv. AN USSR, Technical Cybernetics. (Engl.transl.: Soviet J. Comp. Syst., Sci., **4**, 1979, 79-84. (in Russian).

[4] H. Cramer, *Mathematical methods of statistics*, Princeton University Press. Mathematics, 1999, 575p.

[5] E. Demidenko, *Mixed Models. Theory and Applications with R. Wiley Series in Probability and Statistics*, 1981.

[6] L. Fahrmeir, T. Kneib, S. Lang, B. Marx, *Regression. Models, Methods and Applications*, 2021, Springer.

[7] A.H. Hajiyev, *Linear Regression Models with increasing numbers of unknown Parameters*, Russian Acad. Sci., Doklady, Mathematics, **399(3)**, 2004, 1-5.

[8] A.H. Hajiyev, *Regression Models with Increasing Numbers of Unknown Parameters*, In: International Encyclopedia, Statist. Sciences, Springer, 2011, 1208-1211.

[9] A.H. Hajiyev, *Nonlinear Regression Models with Increasing Numbers of Unknown Parameters*, Russian Acad. Sci., Mathem., **79(3)**, 2009, 339-341.

[10] A.N. Kolmogorov, S.V. Fomin, *Introductory real analysis*, Dover publ. NY, 1975, 393p.

[11] E. Malinvaud, *The consistency of non-linear regressions*, Ann. Mathem. Statist., **41**, 1970, 966-969.

[12] G.F. Seber, C.J. Wild, *Nonlinear regression*, NY, Wiley, 1989.

[13] C.F. Wu, *Asymptotic theory of nonlinear least squares estimation*, Ann. Statist., **9(3)**, 1981, 501-513.

[14] C.F. Wu, *Jackknife, bootstrap, and other resampling methods in regression analysis*, Ann. Statist., **14(4)**, 1986, 1261- 1350.

[15] Sh. Zaks, *Theory of statistical inference*, John Wiley Sons, 1972.

[16] L. Zang, Z. Shi, M.-M. Cheng, Y. Liu, J.W. Biang, J.T. Zhou, Zh.G. Guoyan, Z. Zheng, *Nonlinear Regression via Deep Negative Correlation Learning*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **43(3)**, 2021, 982 - 998.

[17] H. Zhu, Y. Tian, Y. Ren, J. Hu, *A Hybrid Model for Nonlinear Regression with Missing Data Using Quasilinear Kernel*, IEEJ Transactions on Electrical and Electronic Engineering, 15(12), 2020, 1791-1800. https://doi.org/10.1002/tee.23253.

Asaf Hajiyev
Institute of Control Systems, Baku, Azerbaijan
E-mail: asaf.hajiyev@gmail.com

Jiuping Xu
Business School, Sichuan University, Chengdu, China
*E-mail:*icmsem@aliyun.com

Received 14 November 2024
Accepted 30 January 2025