PriMera Scientific
Medicine and Public Health
Volume 6 Issue 6 June 2025
DOI: 10.56831/PSMPH-06-218
ISSN: 2833-5627

# Generalized Regression Models with an Increasing Number of Unknown Parameters

**Type:** Review Article
**Received:** May 16, 2025
**Published:** June 04, 2025

**Citation:**
Asaf Hajiyev., et al. "Generalized Regression Models with an Increasing Number of Unknown Parameters". PriMera Scientific Medicine and Public Health 6.6 (2025): 27-38.



**Asaf Hajiyev[1]* and Jiuping Xu[2]**

[1] *Institute of Control Systems, Ministry of Science and Education, Azerbaijan Republic*

[2] *Business School, Sichuan University, Chengdu, China*

***Corresponding Author:** Asaf Hajiyev, Institute of Control Systems, Ministry of Science and Education, Azerbaijan Republic.

### Abstract

The paper considers the general form of regression models with an increasing number of unknown parameters and different and unknown random error variances. Such models are typical for applications and allow us to solve some practical problems that face difficulties. Linear and nonlinear regression models can be considered as a partial form of this model. The iterated process for calculating the least square estimators for generalized regression models is constructed. The approach for estimating the elements of the covariance matrix of the deviation vector is suggested. Using these results, the method of constructing a confident band for unknown functions in regression models is suggested.

***Keywords:*** generalized regression model; increasing number of unknown parameters; least square estimator; unknown variances of random errors; design matrix; Fisher's matrix; iterated process

## Introduction

The theory of regression analysis has a long history. The earliest form of the method of least squares for regression models was suggested by A. Legendre in 1805 and later by F. Gauss in 1809 in Theoria Motus Corporum Coelestium. Although Legendre formally introduced the method, Gauss had independently developed and applied it as early as 1795 during his astronomical research. However, Gauss's delayed publication, possibly because he did not prioritize formalizing the technique at the time.

C.F. Gauss [12] and A.-M. Legendre [8] independently developed the method of the least squares, and today the academic world considers them as the pioneers of regression theory. Other leading scientists made also essential contributions for the development of regression analysis: U.Yule, K. Pearson [15, 21] C. R. Rao [10], R. Fisher [1] G. Seber [11], P. Huber [7], and others.

The term "regression" was introduced by Francis Galton in the 19th century to describe a biological phenomenon observed in his studies of heredity [3]. Historical accounts differ on when Galton first used the term, with some attributing it to his work in the late 1870s. The phenomenon, now known as "regression toward the mean," describes how the heights of descendants of tall ancestors tend to regress toward the average population. Galton observed this trend while studying the inheritance of traits such as height, using empirical data to quantify the relationship between parent and offspring measurements. While Galton initially applied the concept of regression solely to biological heredity, his pioneering work in quantifying relationships between variables laid the groundwork for its broader statistical applications. Subsequently, Udny Yule and Karl Pearson extended Galton's ideas, generalizing the concept of regression to statistical contexts, including the study of relationships between economic, social, and physical variables. In world literature there are published various types of regression models, for instance, functional regression, spline and quantile regression models, regression models with missing data [14, 16, 17, 23], and others. Below is considered the general form of regression models, and some results for linear and nonlinear regression models can be considered as a special case of generalized regression models.

***Construction of the mathematical model***. Consider the following model

$$y_i = \eta(x_i, \theta) + \varepsilon_i, \ i = \overline{1, N} \qquad (1.1)$$

where $\theta = (\theta_1, \theta_2, \ldots, \theta_m)^T$ - is the vector of unknown parameters. It is supposed that

$\eta(x,\theta)$; $\frac{\partial \eta(x,\theta)}{\partial \theta_i}$; $\frac{\partial^2 \eta(x,\theta)}{\partial \theta_i \cdot \partial \theta_j}$; $i = \overline{1,m}$; $j = \overline{1,m}$- are continuous and bounded

functions, and

$$x \in X \subset R, \theta \in \Theta \subset R^p, 0 < p < N. \qquad (1.2)$$

The families of probability measures on Borel subsets $A$ of the space $X$, are defined as

$$\mu_N(A) = \frac{1}{N}\sum_{i=1}^{N} I_A(x_i),$$

where $I_A(x_i)$ - is the indicator of the set $A$.

It is assumed that $\forall A$ for $N \to \infty$, $\mu_N(A)$ weakly converges to some probability measure $\mu(A)$. Denote such convergence as

$$\mu_N \to \mu \text{ for } N \to \infty \qquad (1.3)$$

We assume also

$$E\varepsilon_i = 0, E\varepsilon_i^2 = \sigma^2(x_i) = \sigma_i \le \sigma_0^2 \qquad (1.4)$$

where $\sigma^2(x)$ - is an unknown bounded function (1.5)

The condition (1.4) is typical for applications and not widely researched in the world literature. There are some papers related to the case of a finite number of unknown parameters and unknown variances of random errors [18] and even for linear and nonlinear regression models with an increasing number of unknown parameters [4-6]. Here generalized model is considered, and the cases of linear and nonlinear models can be considered as a special case of this model. Denote

$$F_N(\theta) \text{ – a matrix with the elements } \frac{\partial \eta(x_i,\theta)}{\partial \theta_j}; i = \overline{1,N}; j = \overline{1,m}.$$

$$\frac{\partial \eta(x,\theta)}{\partial \theta} = \begin{pmatrix} \frac{\partial \eta(x_1,\theta_1,\ldots,\theta_m)}{\partial \theta_1}, \ldots, \frac{\partial \eta(x_1,\theta_1,\ldots,\theta_m)}{\partial \theta_m} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \frac{\partial \eta(x_N,\theta_1,\ldots,\theta_m)}{\partial \theta_1}, \ldots, \frac{\partial \eta(x_N,\theta_1,\ldots,\theta_m)}{\partial \theta_m} \end{pmatrix}$$

$$M_N(\theta) = \left[\frac{F_N^T(\theta)F_N(\theta)}{N}\right] = \frac{1}{N}\int_X \frac{\partial \eta^T(x,\theta)}{\partial \theta} \cdot \frac{\partial \eta(x,\theta)}{\partial \theta}\mu_N(dx)$$

is the normed informatic matrix,

$\lambda_1^{(N)}(\theta), \dots, \lambda_m^{(N)}(\theta)$ - are the eigen values of the matrix $M_N(\theta)$ for which

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

For finding *l.s.e.* for nonlinear regression models, according to [2] we can construct the iterated process which has the following form

$$\theta_N(s+1) = \theta_N(s) + \left(N \cdot M_N(\theta(s))\right)^{-1} F_N^T(\theta(s))\left(y - \eta(x,(\theta(s))\right) \qquad (1.6)$$

Let us investigate the convergence of the process (6). Preliminarily, we remind the following theorem from [9].

***Theorem 1.1***. Assume that the families of functions $G_N(x, \theta)$ which are continuous on $X \times \Theta$ is given where $X$, $\Theta$ - are compact sets.

$$G_N(x,\theta) \underset{\rightarrow}{\rightarrow} G(x,\theta) \text{ (uniformly) for } N \rightarrow \infty.$$

Then

$$\lim_{N\to\infty} \sup_{\Theta} \left| \int_X G_N(x,\theta)\mu_N(dx) - \int_X G(x,\theta)\mu(dx) \right| = 0.$$

From Theorem 1.1, we get

***Corollary 1.1***. Under the conditions (1.2), (1.3), uniformly on $x$ the following expression is held $M_N(\theta) \rightarrow M(\theta)$ for $N \rightarrow \infty$.

***Proof***. In the capacity of $G_N(x, \theta)$ we take

$$G_N(x,\theta) = \frac{F_N^T(\theta)F_N(\theta)}{N}$$

As the elements $G_N(x, \theta)$ are defined as

$$\frac{1}{N}\sum_{i=1}^{N} \frac{\partial \eta^T(x_i,\theta)}{\partial \theta} \cdot \frac{\partial \eta(x_i,\theta)}{\partial \theta} \qquad (1.7)$$

and are continuous and bounded functions according to the condition (1.2), then the series (1.7) converges uniformly for $N \rightarrow \infty$ to some function $G(x, \theta)$. Using the condition (1.2) and applying Theorem 1.1, we get the desired statement.

***Corollary 1.2***. Under the conditions of Corollary 1, uniformly on $\theta$ the following relation is held

$$\frac{\partial M_N(\theta)}{\partial \theta} \rightarrow \frac{\partial M(\theta)}{\partial \theta} \qquad \text{for } N \rightarrow \infty.$$

***Proof***. As

$$\frac{\partial M_N(\theta)}{\partial \theta_k} = \frac{1}{N}\left(\frac{\partial F_N^T(\theta)}{\partial \theta_k} \cdot F_N(\theta) + F_N^T(\theta) \cdot \frac{\partial F_N^T(\theta)}{\partial \theta_k}\right),$$

then considering $\frac{\partial M_N(\theta)}{\partial \theta_k}$, Corollary 1.2 and using Theorem 1.1, we get the desired statement.

***Remark 1.1***. Corollaries 1.1 and 1.2 have proved only the existence of the limits $M_N(\theta)$ and $\frac{\partial M_N(\theta)}{\partial \theta}$, although we do not know the values of these limits.

Consider measurable families $g_{N,y}(x,\theta) \in C(X \times \Theta)$.

Introduce the following conditions:

$$\lim_{N\to\infty} \sup_{X\times\Theta} |g_N(x,\theta) - g(x,\theta)| = 0 \qquad (1.8)$$

$$\sup_{X\times\Theta} |g_{N,y}(x,\theta)| \le C < \infty \qquad (1.9)$$

$$G_N(x,\theta) = E g_{N,y}(x,\theta) \text{ – a continuous function on } X\times\Theta \qquad (1.10)$$

$$\lim_{N\to\infty} \sup_{\Theta} \frac{1}{N} D\left(\sum_{i=1}^{N} g_{N,y_i}(x_i,\theta^*)\right) = 0 \qquad (1.11)$$

Denote

$$S_N(\theta) = \frac{1}{N}\sum_{i=1}^{N} g_{N,y_i}(x_i,\theta), \quad \theta \in B(r); \qquad (1.12)$$

$$G(x,\theta) = \lim_{N\to\infty} G_N(x,\theta) = \lim_{N\to\infty} E g_{N,y}(x,\theta);$$

$$S(\theta) = \int_X G(x,\theta)\mu(dx).$$

***Theorem 1.2*** Assume $\theta \in B(r)$ and the conditions (1.1), (1.6) - (1.9) are held. Then

$$(S_N(\theta) - S(\theta)) \xrightarrow{P} 0, \text{ for } N \to \infty, r \to 0. \qquad (1.13)$$

***Proof***. As $\mu_N(\cdot)$ is a discrete measure, then the ratio (1.12) can be rewritten as

$$S_N(\theta) = \frac{1}{N}\sum_{i=1}^{N} g_{N,y_i}(x_i,\theta) = \int_X g_{N,y}(x,\theta)\mu_N(dx)$$

Then, according to the condition (1.9), we have

$$ES_N(\theta) = \int_X E g_{N,y}(x,\theta)\mu_N(dx) = \int_X G_N(x,\theta)\mu_N(dx) \qquad (1.14)$$

From here, using Theorem 1.1, we get

$$ES_N(\theta) = \int_X G_N(x,\theta)\mu_N(dx) \to \int_X G(x,\theta)\mu(dx), \ \theta \in B(r) \ \text{ for } N \to \infty.$$

Thus, for $N \to \infty$ we have

$$ES_N(\theta) \to \int_X G(x,\theta)\mu(dx) = S(\theta), \theta \in B(r)$$

For $\theta \in B(r)$ taking into consideration (1.12) we have

$$\|S_N(\theta) - ES_N(\theta)\| = \left\| \int_X g_{N,y}(x,\theta)\mu_N(dx) - \int_X G_N(x,\theta)\mu_N(dx) \right\| =$$

$$\left\|\int_X \{g_{N,y}(x,\theta) - G_N(x,\theta) - [g_{N,y}(x,\theta^*) - G_N(x,\theta^*)] + [g_{N,y}(x,\theta^*) - G_N(x,\theta^*)]\}\mu_N(dx)\right\| \le$$

$$\le \left\|\int_X \Delta\varphi(x,\theta,\theta^*)\,\mu_N(dx)\right\| + \left\|\int_X [g_{N,y}(x,\theta^*) - G_N(x,\theta^*)]\mu_N(dx)\right\| \tag{1.15}$$

where $(x, \theta) = g_{N,y}(x, \theta) - G_N(x, \theta)$

$\Delta\varphi(x, \theta, \theta^*) = g_{N,y}(x, \theta) - G_N(x, \theta) - g_{N,y}(x, \theta^*) + G_N(x, \theta^*)$

From ratios (1.11) and (1.12) we have

$$P\{\|S_N(\theta) - ES_N(\theta)\| > a\} - P\left\{\left\|\int_X \Delta\varphi(x,\theta,\theta^*)\,\mu_N(dx)\right\| > a\right\} \le$$

$$\le P\{\|S_N(\theta^*) - ES_N(\theta^*)\| > a\} \tag{1.16}$$

Applying Chebyshev's inequality to the first part of (1.16), we get

$$P\{\|S_N(\theta^*) - ES_N(\theta^*)\| > a\} \le \frac{DS_N(\theta^*)}{a^2} =$$

$$= \frac{\frac{1}{N^2}D\sum_{i=1}^N g_{N,y_i}(x_i,\theta^*)}{a^2} \to 0 \qquad \text{for } N \to \infty$$

Denote $\omega_{N,y}(r)$ - the module of continuity of $\varphi_N(x, \theta)$.

According to the condition (1.6), we have $\omega_{N,y}(r) \le C$.

As $\varphi_N(x, \theta)$ is defined as

$$\varphi_N(x, \theta) = g_{N,y}(x, \theta) - G_N(x, \theta)$$

Then, according to the conditions (1.8) and (1.10), the family of functions $\varphi_N(x, \theta)$ compactly in $X \times \theta$. Then

$$\lim_{r\to 0}\lim_{N\to\infty} \omega_{N,y} = 0$$

almost for all $y$. Thus, all conditions of Lebesgue's theorem about transition under the integral are met. Then we get

$$\lim_{N\to\infty}\lim_{r\to 0} E\omega_{N,y}(r) = 0.$$

Taking into consideration the ratio (1.16), we get

$$(S_N(\theta) - S(\theta)) \xrightarrow{P} 0, \quad \text{for } N \to \infty.$$

Theorem 1.2 has been proved.

Denote

$$\delta_N^* = \delta_N(\theta^*) = y - \eta(x, \theta^*),$$

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

$a_{ij}(\theta); i = \overline{1, N}; j = \overline{1, m}$ - the elements of the matrix $A_N(\theta)$;

$\frac{\partial A_N(\theta)}{\partial \theta}$- is the matrix for the elements of which $\partial a_{ij}\theta$ are held

$$\partial a_{ij}(\theta) = \sum_{k=1}^{m} \frac{\partial a_{ij}(\theta)}{\partial \theta_k};$$

$$\theta = (\theta_1, \theta_2, \dots, \theta_m).$$

***Theorem 1.3*** Suppose $\theta \in B(r)$. If the following condition

$$\frac{\sqrt{m}}{N \cdot \lambda_1^{(N)}(\theta)} \leq C < \infty, for\ N \to \infty, r \to 0 \qquad (1.17)$$

is held then

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

***Proof***. We will use Theorem 2. For this, in the capacity of the family of functions $g_{N,y}(x, \theta)$ we take

$$g_{N,y}(x, \theta) = a_{iN}(x, e) \cdot \delta_N(\theta^*), \qquad i = \overline{1, m};$$

where $a_{iN}(x, e)$ and $(i, N)$ - the elements of the matrix $\frac{\partial A_N(\theta)}{\partial \theta_p}$.

Denote $x_{ij}^{(N)}(\theta); i, j = \overline{1, m}$ - the elements of the matrix $M_N(\theta)^{-1}$.

For $a_{iN}(x, e)$ - we have

$$a_{iN}(\theta) = \sum_{k=1}^{m} x_{ij}^{(N)}(\theta) \cdot f_{KN}(\theta) \leq C_1 \cdot \frac{\sqrt{m}}{\lambda_1^{(N)}(\theta)}, \theta \in B(r)$$

Then from the ratio (1.13) it follows that the conditions of Theorem 1.2 are held. Then for $S_N(\theta)$ we have

$$ES_N(\theta) = \frac{1}{N}\sum_{i=1}^{N} E g_{N,y_i}(x_i, \theta) = \frac{1}{N}\sum_{i=1}^{N} a_{iN} \cdot \delta_N^* = 0.$$

Using Theorem 1.1 we get that for $\theta \in B(r)$

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

Theorem 1.2 has proved.

Denote

$$U_N(\theta) = \theta + A_N(\theta) \cdot \delta_N(\theta);$$

$$L_i^{(N)}(\theta) = \frac{\partial U_N(\theta)}{\partial \theta_i}, \theta \in B(r)$$

$$\tau_N(r) = \max_{1 \leq i \leq m} \sup_{\theta \in B(r)} \left\| L_i^{(N)}(\theta) \right\|.$$

***Theorem 1.4*** Under the conditions 3 the following ratio is held

$$\tau_N(r) \xrightarrow{P} 0 \text{ for } N \to \infty, r \to 0.$$

***Proof***. As

$\left(F_N^T(\theta) \cdot F_N(\theta)\right)^{-1} F_N^T(\theta(s)) \cdot \frac{\partial \delta_N(\theta)}{\partial \theta} = -I_N$ - is the unit matrix then for all $L_i^{(N)}(\theta)$we have

$$L_i^{(N)}(\theta) = \left[I_i + \frac{\partial A_N(\theta)}{\partial \theta_i} \cdot \delta_N(\theta) + A_N(\theta) \cdot \frac{\partial \delta_N(\theta)}{\partial \theta_i}\right] =$$
$$= \left[\frac{\partial A_N(\theta)}{\partial \theta_i} \cdot \delta_N(\theta^*) + \frac{\partial A_N(\theta)}{\partial \theta_i} \cdot \Delta\delta_N(\theta^*, \theta)\right]$$

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

As

$$A_N(\theta) \cdot \frac{\partial \eta_N(x, \theta)}{\partial \theta_i} = -I_i,$$

Then

$$P\{\tau_N(r) > a\} \leq P\left\{\left\|\frac{\partial A_N(\theta)}{\partial \theta_p} \cdot \delta_N(\theta^*)\right\| > a\right\} + P\left\{\left\|\frac{\partial A_N(\theta)}{\partial \theta_p} \cdot \Delta\delta_N(\theta, \theta^*)\right\| > a\right\} \quad (1.18)$$

In (1.18), on the right side, the first term tends to zero $N \to \infty$, $r \to 0$ according to Theorem 1.3, the second term also tends to zero according to Theorem 2 for $N \to \infty$, $r \to 0$. Then we get.

Theorem 1.4 has been proved.

Below is a convergence of the random variables $\varsigma_1$, $\varsigma_2$ will be denoted as

$$\varsigma_1 \to \varsigma_2$$

and it means weak convergence [13].

***Theorem 1.5***. Assume that for $\gamma > 0$ the following ratio is held

$$E|\varepsilon_i|^{2+\gamma} \leq C < \infty$$

and $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N$ - is the sequence of independent random variables.

Then, under the conditions (1.3), we have

$$\sqrt{N} \cdot (\hat{\theta} - \theta^*) \to N(0, \hat{\Sigma}(\theta^*))$$

for $N \to \infty$, where

$$\hat{\Sigma}(\theta^*) = (M(\theta^*))^{-1} \cdot \widetilde{M}(\theta^*) \cdot \left(M(\theta^*)\right)^{-1};$$
$$\widetilde{M}(\theta^*) \lim_{N\to\infty} \widetilde{M}_N(\theta^*) = \lim_{N\to\infty} \frac{F_N^T(\theta^*) \cdot I(\sigma^2) \cdot F_N(\theta^*)}{N};$$

where

$$\mathrm{c}I(\sigma^2) = \begin{pmatrix} \sigma_1^2 & 0 \dots & 0 \\ 0 & \sigma_2^2 \dots & 0 \\ 0 & 0 \dots & \sigma_N^2 \end{pmatrix}$$

(The limit of $\widetilde{M}_N(\theta^*)$ exists according to Corollary 1.1)

***Proof***. Similarly, as it was in the proof of Theorem 1.2 let us introduce nonzero vector $l = (l_1, l_2, \dots, l_m)^T$ of the size $m$.

Then

$$\sqrt{N} \cdot l^T \rho^* = \sqrt{N} \cdot l^T A_N(\theta^*) \cdot \delta_N^* = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} N l^T a_i(\theta^*)(y_i - \eta(x_i, \theta^*))$$

where $a_i(\theta^*)$ $i$-th row of the matrix матрицы $A_N(\theta^*)$.

According to the condition (1.3), we have

$$EN \cdot l^T a_i(\theta^*)\big(y_i - \eta(x_i, \theta^*)\big) = l^T N \cdot E a_i(\theta^*) \cdot \varepsilon_i = 0.$$

Hence, $E\sqrt{N} \cdot l^T A_N(\theta^*) \cdot \delta_N^* = 0$

Consider

$$E\left(\sqrt{N} \cdot l^T A_N(\theta^*) \cdot \delta_N^*\right)^2 = E\sqrt{N} \cdot l^T A_N(\theta^*) \cdot \varepsilon \cdot \varepsilon^T A_N^T(\theta^*) l \sqrt{N} =$$

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

$$= l^T \cdot \hat{\Sigma}(\theta^*) \cdot l \qquad (1.19)$$

As $E\varepsilon_i = E\big(y_i - \eta(x_i, \theta^*)\big)^{2+\gamma} \le C < \infty$, then

$$\frac{C}{\varepsilon} > \frac{1}{\varepsilon \cdot N} \sum_{i=1}^{N} \int |y_i - \eta(x_i, \theta^*)|^{2+\gamma}\, dF_{y_i}(x_i) \qquad (1.20)$$

where $\varepsilon > 0$, $F_{y_i}(x)$ - is the distribution function of $y_i$.

Consider the Lindeberg's fraction

$$\Lambda_N(\varepsilon) = \frac{1}{N} \sum_{i=1}^{N} \int_{|y_i - \eta(x_i, \theta^*)| > \sqrt{N} \cdot \varepsilon} \big(y_i - \eta(x_i, \theta^*)\big)^2\, dF_{y_i}(x_i)$$

We have

$$\frac{C}{\varepsilon} > \frac{1}{\varepsilon \cdot N} \sum_{i=1}^{N} \int |y_i - \eta(x_i, \theta^*)|^{2+\gamma}\, dF_{y_i}(x_i) \ge$$

$$\ge \frac{1}{\varepsilon \cdot N} \sum_{i=1}^{N} \int_{|y_i - \eta(x_i, \theta^*)| \ge \sqrt{N} \cdot \varepsilon} |y_i - \eta(x_i, \theta^*)|^2 \cdot (\varepsilon \cdot \sqrt{N})^\gamma dF_{y_i}(x_i) \ge$$

$$\ge \frac{(\varepsilon \cdot \sqrt{N})^\gamma}{\varepsilon} \cdot \Lambda_N(\varepsilon) \qquad (1.21)$$

Then from (1.17) it follows

$$0 \le \Lambda_N(\varepsilon) \le \frac{C}{\varepsilon^\gamma \cdot N^{\frac{\gamma}{2}}} \quad \text{for } N \to \infty.$$

According to the Central Limit Theorem [9] we have

$$\sqrt{N} \cdot l^T A_N(\theta^*) \cdot \delta_N^* \to N\big(0, l^T \cdot \hat{\Sigma}(\theta^*) \cdot l\big),$$

Using the criterion of normality [22], we have

$$\sqrt{N}\cdot A_N(\theta^*)\cdot\delta_N^*\to N\left(0,\hat{\Sigma}(\theta^*)\right).$$

Theorem 1.5 has been proved.

In the statement of Theorem 1.5, the matrix $\hat{\Sigma}\,(\theta^*)$ - is unknown. Hence, it is necessary to estimate the elements of this matrix.

Suppose $r_B > 0$ and $r_B \to 0$ for $s \to \infty$.

***Theorem 1.6*** If for some $N$ and $r_B$ the following inequality is held

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

Then there exists the random variable $\hat{\theta}N$, such that

$$\left(\theta_N(s)-\hat{\theta}_N\right)\xrightarrow{P}0 \quad \text{for } s\to\infty$$

$$\text{and}\quad \sqrt{N}\left(\hat{\theta}_N-\theta^*\right)\to N(0,\hat{\Sigma}(\theta^*)) \text{ for } N\to\infty.$$

Theorem 1.6 can be considered as a corollary of Theorems 1.4 and 1.5, because all conditions of these theorems are met.

***Estimation of the elements of the covariance matrix for generalized regression models.***

In Theorem 1.5, the matrix $\hat{\Sigma}\,(\theta^*)$ is unknown. Below using $\sqrt{N}$ consistency of the estimator of the unknown parameter $\theta^*$ we will estimate the elements of the matrix $\hat{\Sigma}\,(\theta^*)$. Remind that according to Theorem 1.5, we had

$$\hat{\Sigma}(\theta^*)=(M(\theta^*))^{-1}\cdot\tilde{M}(\theta^*)\cdot\left(M(\theta^*)\right)^{-1}$$

$$\tilde{M}(\theta^*)\lim_{N\to\infty}\tilde{M}_N(\theta^*)=\lim_{N\to\infty}\frac{F_N^T(\theta^*)\cdot I(\sigma^2)\cdot F_N(\theta^*)}{N}$$

Denote $\tilde{m}_{k,l}(\theta)$ - the elements of the matrix $\tilde{M}(\theta^*)$,

$\tilde{m}_{k,l}^{(N)}(\theta)$- are elements of the matrix $\tilde{M}_N(\theta)$

$$\tilde{m}_{k,l}^{(N)}=\frac{1}{N}\left(y-\eta(x,\tilde{\theta})\right)^T\cdot I_{k,l}^2(\tilde{\theta})\cdot(y-\eta(x,\hat{\theta})) \qquad (2.1)$$

$$I_{k,l}^2(\tilde{\theta})=\begin{pmatrix}\varphi_k(x_1,\hat{\theta})\cdot\varphi_l(x_1,\hat{\theta}) & \dots & 0\\ 0 & \varphi_k(x_2,\hat{\theta})\cdot\varphi_l(x_2,\hat{\theta}) & 0\\ 0 & 0 & \varphi_k(x_N,\hat{\theta})\cdot\varphi_l(x_N,\hat{\theta})\end{pmatrix}$$

$0<\hat{\lambda}_1(\theta)\le\dots\le\hat{\lambda}_m(\theta)$- are eigen values of the matrix $M_N(\hat{\theta})$.

***Theorem 2.1*** Assume

$\dfrac{m}{N\cdot\left(\hat{\lambda}_1(\theta)\right)^2}$ is bounded $\forall\theta\in B(r)$

Then

$$E\left(\tilde{m}_{k,l}^{(N)}-\tilde{m}_{k,l}^{(N)}(\theta^*)\right)\to 0,$$

$$\left(\tilde{m}_{k,l}^{(N)}-\tilde{m}_{k,l}^{(N)}(\theta^*)\right)\xrightarrow{P}0 \text{for } N\to\infty.$$

***Proof***. Consider

$$y-\eta\left(x,\hat{\theta}\right)=\eta(x,\theta^*)+\varepsilon-\eta\left(x,\hat{\theta}\right)=\frac{\partial\eta(x,\theta)}{\partial\theta}\left(\theta^*-\hat{\theta}\right)+\varepsilon+o(r)=$$

$$=-F_N\left(\hat{\theta}\right)\left(\hat{\theta}-\theta^*\right)+\varepsilon+o(r). \qquad (2.2)$$

For $\theta$ ($s$ + 1) ∈ $B(r)$ from the ratio (2.2) we have

$$\theta(s+1)-\theta^{*}=\theta(s)+\left[F_N^T(\theta(s))\cdot F_N(\theta(s))\right]^{-1}F_N^T(\theta(s))\times$$
$$\times\left(y-\eta(x,\theta(s))\right)-\theta^{*}=\theta(s)+\left[F_N^T(\theta(s))\cdot F_N(\theta(s))\right]^{-1}F_N^T(\theta(s))\times$$
$$\times\left(\eta(x,\theta^{*})+\varepsilon-\eta(x,\theta(s))\right)-\theta^{*}=$$
$$=\theta(s)+\left[F_N^T(\theta(s))\cdot F_N(\theta(s))\right]^{-1}F_N^T(\theta(s))\cdot\left[(\theta^{*}-\theta(s))+\varepsilon+o(r)\right]-\theta^{*}=$$
$$=+\left[F_N^T(\theta(s))\cdot F_N(\theta(s))\right]^{-1}F_N^T(\theta(s))\cdot\varepsilon+o(r).$$

As the elements of the matrix $F_N(\theta)$ are continuous, then for $s\to\infty$, $r\to 0$ then, according to Theorem 1.6 we have

$$\left[(\hat{\theta}-\theta^{*})-\left(F_N^T(\hat{\theta})\cdot F_N(\hat{\theta})\right)^{-1}\cdot F_N^T(\hat{\theta})\cdot\varepsilon\right]\xrightarrow{P}0$$

Then from (2.2) it follows

$$y-\eta(x,\hat{\theta})=\varepsilon-F_N(\hat{\theta})\cdot\left(F_N^T(\hat{\theta})\cdot F_N(\hat{\theta})\right)^{-1}\cdot F_N^T(\hat{\theta})\cdot\varepsilon=$$
$$=\left[I-F_N(\hat{\theta})\cdot\left(F_N^T(\hat{\theta})\cdot F_N(\hat{\theta})\right)^{-1}\cdot F_N^T(\hat{\theta})\right]\cdot\varepsilon \qquad (2.3)$$

Substituting (2.3) in (2.2) and opening the parenthesis, we have

$$\frac{1}{N}\left(y-\eta(x,\hat{\theta})\right)^T\cdot I_{k,l}^{(2)}(\hat{\theta})\cdot\left(y-\eta(x,\hat{\theta})\right)=\frac{\varepsilon^T\cdot I_{k,l}^{(2)}(\hat{\theta})\cdot\varepsilon}{N}-$$
$$-\frac{\varepsilon^T\cdot I_{k,l}^{(2)}(\hat{\theta})\cdot F_N(\hat{\theta})\cdot\left(F_N^T(\hat{\theta})\cdot F_N(\hat{\theta})\right)\cdot\varepsilon}{N}-\varepsilon^T\cdot F_N(\hat{\theta})\times$$
$$\times\frac{\left(F_N^T(\hat{\theta})\cdot F_N(\hat{\theta})\right)^{-1}\cdot F_N^T(\hat{\theta})\cdot I_{k,l}(\hat{\theta})\cdot\varepsilon}{N}+\varepsilon^T\cdot F_N(\hat{\theta})\times$$
$$\times\frac{\left(F_N^T(\hat{\theta})\cdot F_N(\hat{\theta})\right)^{-1}\cdot F_N^T(\hat{\theta})\cdot I_{k,l}(\hat{\theta})\cdot F_N(\hat{\theta})\left(F_N^T(\hat{\theta})\cdot F_N(\hat{\theta})\right)^{-1}\cdot F_N^T(\hat{\theta})\cdot\varepsilon}{N} \qquad (2.4)$$

Repeating the way of Theorem 2.1, which proves (2.4) and tendency to zero of the last terms on probability for $N\to\infty$, uniformly on $x$. Then we get that $\hat{m}_{k,l}^{(N)}(\hat{\theta})$ is asymptotically unbiased estimator of the element $\tilde{m}_{k,l}(\theta^{*})$ and according to the same Theorem 2.1 it is also consistent estimator.

Using Corollary 1.1, we get that for the matrix $\tilde{M}_N(\hat{\theta})$ there exists the limit matrix $\tilde{M}$, and the elements of the matrix $\tilde{M}_N(\hat{\theta})$ for $N\to\infty$ are asymptotically unbiased and consistent estimators corresponding elements of the matrix $M(\theta^*)$.

Consider the matrix

$$\bar{B}_N=\frac{1}{N}\left[\frac{F_N^T(\hat{\theta})\cdot F_N(\hat{\theta})}{N}\right]^{-1}\cdot\hat{B}\cdot\left[\frac{F_N^T(\hat{\theta})\cdot F_N(\hat{\theta})}{N}\right]^{-1}$$

Denote $\bar{b}_{k,l}$ - the elements of the matrix $\bar{B}_N$, and $c_{k,l}$ - the elements of the matrix $\hat{\Sigma}(\theta^{*})$.

Then

$$\bar{b}_{k,l}=\frac{1}{N}\sum_{i_1=1}^{m}\sum_{i_2=1}^{m}x_{kl_1}\cdot\bar{b}_{i_1i_2}\cdot x_{i_2l}\,,$$
$$\tilde{c}_{k,l}=\frac{1}{N}\sum_{i_1=1}^{m}\sum_{i_2=1}^{m}x_{kl_1}\cdot\tilde{b}_{i_1i_2}\cdot x_{i_2l}\,.$$

Under the conditions of Theorem 1.6 we have

$$\left|\bar{b}_{k,l} - \tilde{c}_{k,l}\right| = \frac{1}{N}\left|\sum_{i_1=1}^{m}\sum_{i_2=1}^{m} x_{kl_1}\cdot\left(\bar{b}_{i_1 i_2} - \tilde{b}_{i_1 i_2}\right)\cdot x_{i_2 l}\right| \le$$
$$\le \frac{m}{\left(\hat{\lambda}_1(N)\right)^2\cdot N}\max\left|\bar{b}_{i_1 i_2} - \tilde{b}_{i_1 i_2}\right| \to 0$$

which proves Theorem 2.1.

For constructing a confidence band for the function $\eta(x,\theta)$ we can use Theorems 1.5 and 1.6 and the approach suggested in [4].

***Remark 2.1***. It is necessary to underline the advantages and flaws of the suggested method above. The advantage of this approach is that the method allows for investigating the regression models where the errors of random variables have different and unknown variances, and at one point of observation, there is no more than one response. Such models are typical for applications and have not been widely research.

The deficiency of this approach is that all results (estimation of unknown parameters and the elements of a covariance matrix) have an asymptotic structure and have an efficiency for a large number of observations.

## Conclusion

The mathematical model of generalized regression models with an increasing number of unknown parameters is suggested. Linear and nonlinear regression models can be considered as a partial form of this model. The iterated process for calculating the least square estimators is constructed. A new method for estimating the elements of a covariance matrix has been suggested. Using these results, the method of constructing a confidence band for an unknown function in regression models is suggested.

## References


1. Aldrich J. “Fisher and Regression”. Statist. Sci 20.4 (2005): 401-417.
2. Demidenko E. “Mixed Models: Theory and Applications with R”. Wiley Series in Probability and Statistics (1981).
3. Galton F. “Typical laws of heredity”. Nature 15 (1877): 492-495, 512- 514, 532-533.
4. Hajiyev AH. “Regression Models with Increasing Numbers of Unknown Parameters”. In: International Encyclopedia. Statist. Sciences, Springer (2011): 1208-1211.
5. Hajiyev AH. “Nonlinear Regression Models with Increasing Numbers of Unknown Parameters”. Russian Acad. Sci., Mathem 79.3 (2009): 339-341.
6. Hajiyev AH. “Linear Regression Models with increasing numbers of unknown Parameters”. Russian Acad.Sci., Doklady, Mathematics 399.3 (2004): 1-5.
7. Huber PJ. “Robust regression: Asymptotics, conjectures and Monte Carlo”. The Annals of Statistics 1 (1973): 799-821.
8. Legendre A.-M On Least Squares. Translated from the French by Professor Henry A Ruger and Professor Helen M Walker, Teachers College, Columbia University, New York City (1805).
9. Prohorov YuV and Rozanov YuV. “Probability theory. Basic concepts-limit theorems-random processes”. Springer (1969).
10. Rao CR. “Linear Statistical Inference and its Applications”. John Wiley Series in Probability and Statistics (1973).
11. Seber GF. “Linear Regression Analysis”. NY, Wiley Ser. Probab. and Statistics (2003).
12. Sheinin O. “Gauss and the method least squares”. Jahrbucher fur Nationalokonomie und statistic 219.3+4 (1999): 458-467.
13. Shiryaev AN. Probability. Springer-Verlag, New York-Berlin- Heidelberg (2013): 580.
14. Sun Q, Zhou WX and Fan J. “Adaptive Huber Regression”. Am Stat. Assoc 115.529 (2020): 254-265.
15. Stanton JM. “Galton, Pearson, and the Peas: A brief history of linear regression for statistics instructors”. Journal of Statist. Edu-

cation 9.3 (2001).
16. Wand MP. “Comparison of Regression Spline Smoothing Procedures”. Computational Statistics 15 (2000): 443-462.
17. Wang L, Wu Y and Li R. “Quantile regression for analyzing heterogeneity in ultra-high dimension”. Journal of the American Statistical Association 107 (2012): 214-222.
18. Wu CF. “Jacknife, bootstrap, and other resampling methods in regression analysis”. Ann. Statist 14.4 (1986): 1261-1350.
19. Wu CFJ and Hamada MS. “Experiment. Planning, Analysis and Optimization”. Wiley, Ser. Probability and Statistics (2009).
20. Yao F, Muller H-G and Wang J-L. “Functional linear regression analysis for longitudinal data”. Ann. Statist 33.6 (2005): 2873-2903.
21. Yule GU and Pearson K. “On theory of consistence of logical class- frequences, and its geometrical representation”. “Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences (1997): 287-299.
22. Zaks Sh. “Theory of statistical inference”. John Wiley Sons (1972).
23. Zapf A, Wiessner C and König IR. “Regression Analysis and Their Particularities in Observational Studies”. Dtsch Arztebl Int 121.4 (2024): 128-134.