Journal of Economics
and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq
DOI: https://doi.org/10.33095/jeas.v28i134.2428

Estimating Parameters via L-Linear Method for Second-Order Regression of Polynomial Model

Mahdi Younis Mohammad
Department of Statistics,
College of Administration and Economics, University of Baghdad
Baghdad, Iraq
Mehdiiraqi62@gmail.com

Ahmed Dheyab Ahmed
Department of Statistics,
College of Administration and Economics, University of Baghdad
Baghdad, Iraq
ahmedthieb19@gmail.com

Received:6/10/2022 **Accepted:30/10/2022** **Published: December / 2022**

Abstract

In this paper, estimation of the parameters of the second-order the polynomial regression model was obtained

when the error is distributed it is distribution long-tailed symmetric because the presence of outliers and using Maximum Likelihood, and L-Method. simulation is illustrated, and Mean Squared Error (MSE) used an evaluation criterion. This research includes six sample sizes (50,60,80,90,100,120) and includes the application of the methods the medical data representing diabetes.

Paper type Research paper.

Keywords: Regression of Polynomial, long-tailed Symmetric distribution, Maximum Likelihood estimation Method, L-Method.

1. Introduction

Regression (Al-Mashhadani, 1989) is a statistical research technique used to find a relationship between two or more variables. This relationship is approximated by a line, where regression analysis is used to identify the best line that fits the existing data. The relationship between the variables is also expressed in a mathematical relationship through which the effect of the explanatory variables on the response variable can be estimated. The Regression analysis has importance because it allows the researcher to test the relationship between two or more variables to predict the future. The linear regression is considered a first-degree model, and when there is a single explanatory variable, this is called the Simple Linear Regression, while it will be called a Multiple Linear Regression if there are several explanatory variables. But non-linear regression is used when the variables have an exponent greater than one such as a polynomial. Waebe (2008) concluded that the financial data represents either losses or returns and frequently has skewness in their behavior, researchers Domowitz and white (Domowitz and White , 1984) used the nonlinear least squares method if there is a problem of heterogeneity of variances of random error

2. Material and Methods

2.1. Polynomial Regression

In polynomial regression (Ostertagova,2012) , the relationship between the explanatory variables X and the dependent variable Y is modeled, provided that X is of degree n, where regression of polynomial (Islam, 2009) is a type of non-linear regression (Akkaya, 2008) and it is applications in economic, medical and others.

The form of the polynomial regression model of degree k is:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x^2_{i2} + \ldots + \beta_k x^k_k + e_i \qquad \ldots (1)$$

(k) is the degree of the model OLS estimators are not effective in for estimating the parameters because of the presence of outliers (Filzmoser , 2004) , (Al-Mashhadani , 1989) or extreme values that could be the source of an error in reading or recording the data, or the source of the outliers may be from the statistical population, so there for researchers, Akkaya and Tiku (2008) the Multiple Linear Regression model as follows[1] :

$$y_i = \theta_0 + \sum_{j=1}^{q} \theta_j u_{ij} + e_i \qquad 1 \le j \le q, 1 \le i \le n \qquad \ldots (2)$$

$$u_{ij} = \frac{x_{ij} - \bar{x}_j}{s_i}$$

$$\bar{x}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}$$

$$s_j^2 = \sum_{i=1}^{n} \frac{1}{n} (x_{ij} - \bar{x}_j)^2$$

Koenker and Park (Koenker and Park , 1996) describe the algorithm for calculating quantile regression in event that the independent variables have nonlinear functions. In 1995, researcher Al-Sulaiman (Al-Sulaiman ,1995) presented an estimate of two nonlinear regression models; the first is the Logistic model, and the second is the probit model, and it uses the least squares method to estimate the multiple case of linear regression model and the maximum likelihood method to estimate the nonlinear regression models. In 2008, Akkaya and Tiku estimated a polynomial regression model assuming that the error distribution is LTS ,Generalized Logistic and short-tiled symmetric and the estimation was applied to three chemical experiments.

The Second Order Regression of the Polynomial Model has the following equation:

$$y_i = \theta_0 + \sum_{j=1}^{q} \theta_j u_{ij} + \sum_{j=1}^{q} \theta_{jj} u_{ij}^2 + \sum_{j=1}^{q-1} \sum_{k=j+1}^{q} \theta_{jk} u_{ij} u_{ik} + e_i \qquad \dots (3)$$

$$1 \le j \le q, \qquad 1 \le i \le n$$

q: *Number of explanatory variables*, **model 3 can be written in matrix form as follows:**

$$Y = U\theta + e \qquad \dots (4)$$

where:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{bmatrix} \quad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ \vdots \\ e_n \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_0 \\ \vdots \\ \theta_q \\ \theta_{11} \\ \vdots \\ \theta_{qq} \\ \theta_{12} \\ \vdots \\ \theta_{q-1,q} \end{bmatrix},$$

$$U = \begin{pmatrix} 1 & u_{11} & \dots & u_{1q} & u_{11}^2 & \dots & u_{1q}^2 & u_{11}u_{12} & \dots & u_{1q-1}u_{1q} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & u_{n1} & \dots & u_{nq} & u_{n1}^2 & \dots & u_{nq}^2 & u_{n1}u_{n2} & \dots & u_{n,q-1}u_{nq} \end{pmatrix}$$

2.2 Methods of Estimation

The parameters of the second-order polynomial case of the regression model can be estimated by the maximum likelihood method and L–method.

The Ordinary Least Square Method formula is:

$$\tilde{\theta} = (\grave{U}U)^{\grave{}-1}(UY) \qquad \dots (5)$$

$$\sigma^2 = \frac{Ee^2}{n-c} = \frac{(Y-U\theta)'(Y-U\theta)}{n-c},$$

$$c = 1 + 2q + \frac{q(q-1)}{2}$$

2.2.1 Maximum Likelihood Method

We use the Maximum Likelihood method to estimate the regression parameters, assuming that the error is distributed Long Tailed symmetric, is as follows (Domowitz and White,1984) (Phan, and Tan , 2011) :

$$f(e) = \frac{\Gamma(p)}{\sigma\sqrt{k}\,\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{p-1}{2}\right)}\left\{1+\frac{e^2}{k\sigma^2}\right\}^{-p} . -\infty < e < +\infty \qquad \dots (6)$$

$$E(e) = 0 .\ \ V(e) = \sigma^2, \qquad k = 2p-3, \quad t = \sqrt{\frac{v}{k}}\,\frac{e}{\sigma}$$

σ: represents the scale parameter, and p: the shape parameter, and their likelihood function can be as follows:

$$L = \prod_{i=1}^{n} \frac{\Gamma(p)}{\sigma\sqrt{k}\,\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{p-1}{2}\right)}\left\{1+\frac{e^2}{k\sigma^2}\right\}^{-p}$$

$$\ln L = n\ln d - n\ln\sigma - p\sum_{i=1}^{n}\ln\left(1+\frac{{z_i}^2}{k}\right) \qquad \dots (7)$$

$$z_i = \frac{e_i}{\sigma}, \quad d = \frac{\Gamma(p)}{\sqrt{k}\,\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{p-1}{2}\right)}$$

$$e_i = y_i - \left(\theta_0 + \sum_{j=1}^{q}\theta_j\,u_{ij} + \sum_{j=1}^{q}\theta_{jj}\,u_{ij}^2 + \sum_{j=1}^{q-1}\sum_{k=j+1}^{q}\theta_{jk}\,u_{ij}\,u_{ik}\right)$$

$$\frac{\partial L}{\partial\theta_0} = \frac{2p}{k\sigma}\sum_{i=1}^{n}\frac{z_i}{1+\frac{z_i^2}{k}} = 0 \qquad \dots (8)$$

$$\frac{\partial L}{\partial\theta_j} = \frac{2p}{k\sigma}\sum_{i=1}^{n}\sum_{j=1}^{q}\frac{u_{ij}z_i}{1+\frac{z_i^2}{k}} = 0 \qquad \dots (9)$$

$$\frac{\partial L}{\partial\theta_{jj}} = \frac{2p}{k\sigma}\sum_{i=1}^{n}\sum_{j=1}^{q}\frac{{u_{ij}}^2 z_i}{1+\frac{z_i^2}{k}} = 0 \qquad \dots (10)$$

$$\frac{\partial L}{\partial\theta_{jk}} = \frac{2p}{k\sigma}\sum_{i=1}^{n}\sum_{j=1}^{q-1}\sum_{k=j+1}^{q}\frac{u_{ij}u_{ik}z_i}{1+\frac{z_i^2}{k}} = 0 \qquad \dots (11)$$

$$\frac{\partial L}{\partial\sigma} = \frac{2p}{k\sigma}\sum_{i=1}^{n}\frac{{z_i}^2}{1+\frac{z_i^2}{k}} = 0 \qquad \dots (12)$$

It is not possible to find solutions to these equations because the equations are non-linear, so it requires finding a technique to solve the equations numerically, so we will use one of the numerical methods, which is the Newton–Raphson method to find an estimate of the model parameters, which depends on linear approximation technique (Al-Sulaiman , 1995) (Jun Sun and Xiao-Jun , 2012) .

2.2.2 L- Method

It is one of the estimation methods that depend on the minimum sum of absolute errors (MSAE) [12], and that is by replacing the squares of errors with absolute values. The least absolute value estimates (LAV) can be obtained as follows (Jun Sun and Xiao-Jun , 2012) (Koenker and Park , 1996) :

$$\hat{\beta}_{LA} = \arg\min_{\beta} \sum_{i=1}^{n} |e_i(\theta)| \qquad \dots (13)$$

where the error is calculated by the following equation:

$e_i \ (\theta) = y_i - \hat{y}. i = 1. \dots \dots . n \ )$

To estimator $\hat{\beta}$, the minimization problem is reformulated by a linear approximation of the nonlinear regression, and the necessary condition for a vector $d \in [-1.1]^n$ that helps to solve a series of problems to minimize errors, where:

$d_i = sign(e_i) \qquad if \ e_i \neq 0$

$d_i \in (-1.1) \qquad if \ e_i = 0$

(-1.1)open period between-1and 1

where (e_i) : the sign of error to solve the problem L for each iteration. By applying the Meketon algorithm, the formula for estimating the parameters is: (Koenker and Bassett , 1978)

$$\theta = \left(\dot{U}' D^2 U\right)^{-1} \dot{U}' D^2 \, U \qquad \dots (14)$$

D is a diagonal matrix with the following formula:

$$D = diag(min\{(1 + d_i).(1 - d_i)\}) \qquad \dots (15)$$

3. Discussion of Results

3.1 Simulation

In the simulation, the shape and scale parameters are P=3,4,5,7 and ($\sigma^2 = 1$). As for the values of the parameters of the regression model, which are taken from an applied experiment, they are as follows:

Par.	θ_0	θ_1	θ_2	θ_3	θ_{11}	θ_{22}	θ_{33}	θ_{12}	θ_{13}	θ_{23}
value	77.21	-8.79	-7.43	-0.05	-3.06	-3.52	-1.73	-4.68	-2.08	-1.17

The distribution of generating values for the random variable is long-tailed symmetric as follows:

$$e = \frac{t\sigma}{\sqrt{\frac{v}{k}}} \qquad \dots (16)$$

t represents the t-distribution random data generated by Matlab program using a command (trad)

$t = trad(2p - 1)$

Six sample sizes were selected (50, 60, 80, 90, 100,120) and the experiment was repeated 1000 times. To compare the estimation methods, the statistical scale MSE was used and the program was written in Matlab program. The following are the simulation results according to Table 1:

Table 1. MSE of Model and L-Estimator

p	3	4	5	7
n	L	L	L	L
50	2.907358	2.614178	2.395216	2.284943
60	2.8378	2.540947	2.239245	2.109064
80	2.760688	2.34735	2.127075	1.957603
90	2.781875	2.278346	2.046172	1.91214
100	2.652992	2.306338	1.994387	1.891824
120	2.658073	2.173926	1.968745	1.773763

The results of the Maximum Likelihood Method (MLM) were neglected by using simulation because it did not give normal results, and the Robust L method gave the least mean squared error for the model and its value decreases when the sample size increases and when the P value increases.

3.2 Real Data Application

The data were collected and gathered from a laboratory belonging to the Iraqi Ministry of Health: MOH (Gilgamesh Laboratory, 2022), where they represent the evidence of the level of sugar in the body, with 90 individuals as a sample size understudy. The variables represent the following: c-peptide, X1: HBa1c, X2: RBS, X3: UREA: y

After taking the standard degree of the dependent variable Y,the Kolmogorov-Smirnov test was used, and the data were processed and then, they were tested again. The main finding was that long-tailed symmetric occurred to be their distribution. The computed value D = 0.1411, which is less than the table values at the level of significance $K_{0.05} = 0.1434$ and$K_{0.01} = 0.1718$, this means accepting the null hypothesis H_0, that is, the data have long tailed symmetric distribution.

To estimate the parameters of the second-order multivariate regression model, the L-Estimator was used and the results were as in the following table:

Table 2. Estimated values of model parameters when p = 7

parameters	θ_0	θ_1	θ_2	θ_3	θ_{11}
Value	-0.07069	0.020731	-0.17927	-0.16687	0.055527
parameters	θ_{22}	θ_{33}	θ_{12}	θ_{13}	θ_{23}
Value	0.062	0.188257	-0.06102	-0.57748	0.271651

$$\hat{y} = -0.07069 + 0.020731u_{i1} - 0.17927u_{i2} - 0.16687u_{i3} + 0.055527u_{i1}^2 + 0.062u_{i2}^2 + 0.188257u_{i3}^2 - 0.06102u_{i1}u_{i2} - 0.57748u_{i1}u_{i3} + 0.271651u_{i2}$$

4. Conclusion

1. The L - method showed its superiority for parameters' estimation of the polynomial regression model.

2. The best values for the parameters of shape and size are ($p = 7, \sigma = 1$)

3. At a value of ($p = 7$)and for all sample sizes, the L – method has a lower MSE of the model.

4. The associated relationship between the C-PeP variable and the variable RBC occurred to be positive. Conversely,a negative relationship occurred between the variable C-PeP and the HBAIC variable, and the C-PeP variable and the UREA variable, too.

References

1.Akkaya, A.D., and Tiku, M.L. (2008) " Robust estimation in multiple linear regression model with non-Gaussian noise". Automatica 44, 407-417

2. Al-Mashhadani, M. H. & Hormuz, A. H. (1989). Statistics. College of Administration and Economics, University of Baghdad.

3. Al-Sulaiman, M. S. (1995). Using Linear and Nonlinear Models in Estimation. Master's Thesis, College of Administration and Economics - University of Baghdad.

4. Domowitz, I., and White, B. H. (1984). nonlinear regression with dependent observations",Econometrica,vol.52.pp(143-161).

5. El-Attar, R.A., M. Vidyasagar, and S.R.K. Dutta, 1979, An algorithm for minimization with application to nonlinear approximation. SIAM Journal of Numerical Analysis 16, 70-86

6. Filzmoser, P.(2004), " A multivariate Outlier detection Method", Department of statistics and Probability Theory. Vienna, Austria, Volume 1,PP. 18-22

7. Islam,T., Shaibur,M.R. and Hossain,S.S.(2009)." Effectivity of Modified Maximum Likelihood Estimators Using Selected Ranked Set Sampling Data". AUSTRIAN JOURNAL OF Statistics Volume 38, Number 2, 109–120

8. Jun Sun, Choi-Hong Lai and Xiao-Jun Wu, (2012). Particle Swarm Optimization: Classical and Quantum Perspectives, Chapman and Hall/CRC: Numerical analysis and scientific computing, International Standard Book Number-13: 978-1-4398-3577-7, ISBN : 978-L-4398-357b-0 ,Pp. 77. 78. 95-97.

9. Koenker and Park . J, 1996 ,An interior point algorithm for nonlinear quantile regression,Universite of Llliois.

10. Koenker, R. and V. d'Orey, 1987, Computing regression quantiles, Applied Statistics 36, 383-390

11. Koenker, R.W. and Bassett, G.W., 1978, Regression quantiles, Econometrica 46, 33-50

12. Meketon, M., 1986, Least absolute value regression, technical report (AT&T Bell Laboratories, Holmdel, NJ)

13. Ostertagova, E.(2012)"Modelling using polynomial regression" Procedia Engineering 48 500 – 506

14. Phan, Q., H., and Tan, S. L. (2011)." Mitigation of GPS periodic Multpath Using non Linear Rgression" 19th European signal processing conference (Eusipco 2011)

تقدير المعلمات باستعمال طريقة L الخطية لانموذج الانحدار المتعدد الحدود من الدرجة الثانية

الباحث/ مهدي يونس محمد
قسم الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد
بغداد، العراق
Mehdiiraqi62@gmail.com

أ.د. احمد ذياب احمد
قسم الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد
بغداد، العراق
ahmedthieb19@gmail.com

Received:6/10/2022 Accepted:30/10/2022 Published: December / 2022

مستخلص البحث

في البحث يتم الحصول على تقديرات معلمات انموذج الانحدار متعدد العوامل متعدد الحدود من الدرجة الثانية عندما يتوزع الخطأ العشوائي طويل الذيل المتماثل بسبب وجود القيم الشاذة وباستعمال طريقة الإمكان الأعظم وطريقة L وبأسلوب المحاكاة وباستعمال ست احجام للعينات (50، 60، 80، 90، 100، 120) وتطبيق الطرائق على بيانات طبية تمثل مرض السكر.

نوع البحث: ورقة بحثية

المصطلحات الرئيسة للبحث: الانحدار المتعدد الحدود, توزيع طويل الذيل المتماثل, طريقة الامكان الاعظم, طريقة L, المحاكاة.

*البحث مستل من رسالة ماجستير