

On an Alternative Method of Estimation of Polynomial Regression Models

¹O.D. Ogunwale, ²O.Y. Halid and ³J. Sunday

^{1,2}Department of Mathematical Sciences, University of Ado-Ekiti, Ekiti State, Nigeria.
 Department of Mathematical Sciences, Adamawa State University, Mubi Nigeria.

Abstract: The OLS and ML methods had been important tools of obtaining estimates of parameters in regression analysis. This paper proposed a numerical analytic method as an alternative approach to the OLS method for a Polynomial Regression Model of the third degree. The model was fitted to a data of output and total production cost of a commodity using the OLS and the computational group average (of numerical analysis) techniques. We obtained very close parameter estimates in both cases.

Key words: *OLS-Ordinary Least Square, ML-maximum likelihood, Regression, Parameters, Computational Group Average*

INTRODUCTION

Statistical models have numerous real life applications in all professional walks of life. One of them is found in econometrics meaning econometric measurements. The models according to functional forms are categorized into linear and non-linear models. The polynomial regression models are important class of non-linear models which have extensive use in econometric researches especially situations relating to cost and production. Its non-linearity came from the fact that the relationship between marginal cost (MC) of production(Y) of a commodity and its output (X) is non-linear since a U-shaped curve (parabola) is usually observed in its scattered diagram. This of course, represents a quadratic form. That is, a polynomial regression model of degree two whose quadratic form is given by

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i, \quad i = 1, \dots, n \tag{1}$$

X_i is assumed to be fixed called the independent variable and consequently X_i^2 is also fixed while ε_i is

the random error term. β_0, β_1 , and β_2 are the coefficients of the equation. Its general form is,

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \dots + \beta_k X_i^k + \varepsilon_i \tag{2}$$

called the polynomial regression model of degree k.

The exponential, logistic and Gompertz models are also important class of non-linear regression models with great applications in growth and population studies having respective forms;

$$Y_i = \beta_1 e^{\beta_2 X_i} + \varepsilon_i \tag{3}$$

$$Y_i = \frac{\beta_1}{1 + e^{\beta_2 X_i}} + \varepsilon_i \tag{4}$$

$$Y_i = \beta_1 e^{-\beta_2 X_i} + \varepsilon_i \tag{5}$$

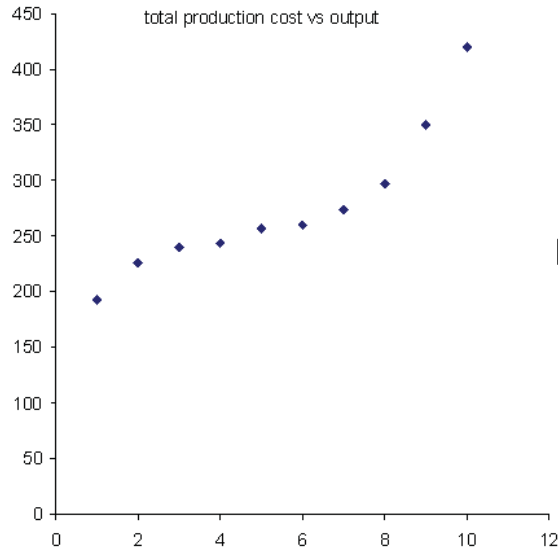
Corresponding Author: O.D. Ogunwale, Department of Mathematical Sciences, University of Ado-Ekiti, Ekiti State, Nigeria.

MATERIALS AND METHODS

A data of output and total cost of production of a commodity in the short run is considered below

Output	1	2	3	4	5	6	7	8	9	10
TPC	193	226	240	244	257	260	274	297	350	420

The scattered diagram is given by the succeeding graph



Scattered diagram of output and total production cost

We observed an S-shaped curve which follows the principle of the celebrated law of diminishing returns in the short run. This is identical to a cubic function.

We therefore propose a polynomial regression model of the third degree for the data, so that

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \varepsilon_i \tag{6}$$

The OLS Method:

The ordinary least square method of estimation in regression analysis is used as follows since the model does not violate the non multicollinearity assumption.

Then,

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i - \beta_2 X_i^2 - \beta_3 X_i^3)^2 \tag{7}$$

taking the partial derivatives of equation (7) w.r.t. the parameters β_0 , β_1 , β_2 , and β_3 and setting the resulting results to zero. These yield;

$$\frac{\partial \sum \varepsilon_i^2}{\partial \beta_0} = (-2) \sum (Y_i - \beta_0 - \beta_1 X_i - \beta_2 X_i^2 - \beta_3 X_i^3) = 0 \tag{8}$$

$$\frac{\partial \sum \varepsilon_i^2}{\partial \beta_1} = (-2 X_i) \sum (Y_i - \beta_0 - \beta_1 X_i - \beta_2 X_i^2 - \beta_3 X_i^3) = 0 \tag{9}$$

$$\frac{\partial \sum \varepsilon_i^2}{\partial \beta_2} = (-2X_i^2) \sum (Y_i - \beta_0 - \beta_1 X_i - \beta_2 X_i^2 - \beta_3 X_i^3) = 0 \quad (10)$$

$$\frac{\partial \sum \varepsilon_i^2}{\partial \beta_3} = (-2X_i^3) \sum (Y_i - \beta_0 - \beta_1 X_i - \beta_2 X_i^2 - \beta_3 X_i^3) = 0 \quad (11)$$

We note that equations (8) - (11) can also be written as

$$\sum \varepsilon_i = 0 \quad (8a)$$

$$\sum X_i \varepsilon_i = 0 \quad (9a)$$

$$\sum X_i^2 \varepsilon_i = 0 \quad (10a)$$

$$\sum X_i^3 \varepsilon_i = 0 \quad (11a)$$

Equations (8a) – (11a) show that the properties of least squares fit namely that the residual is equal to zero and they are uncorrelated with the independent variables.

We solve (8) - (11) further to obtain their normal equations given by equations (8b) – (11b) so that;

$$\sum Y_i = n\beta_0 + \beta_1 \sum X_i + \beta_2 \sum X_i^2 + \beta_3 \sum X_i^3 \quad (8b)$$

$$\sum X_i Y_i = \beta_0 \sum X_i + \beta_1 \sum X_i^2 + \beta_2 \sum X_i^3 + \beta_3 \sum X_i^4 \quad (9b)$$

$$\sum X_i^2 Y_i = \beta_0 \sum X_i^2 + \beta_1 \sum X_i^3 + \beta_2 \sum X_i^4 + \beta_3 \sum X_i^5 \quad (10b)$$

$$\sum X_i^3 Y_i = \beta_0 \sum X_i^3 + \beta_1 \sum X_i^4 + \beta_2 \sum X_i^5 + \beta_3 \sum X_i^6 \quad (11b)$$

Fitting the data of output and total production cost to equations (8b) – (11b) and solving for the unknown parameters, we obtain the OLS estimates

$$\hat{\beta}_0 = 141.7667, \hat{\beta}_1 = 63.4776, \hat{\beta}_2 = -12.9615, \text{ and } \hat{\beta}_3 = 0.9396$$

Hence, we have

$$Y = 141.7667 + 63.4776X - 12.9615X^2 + 0.9396X^3 \quad (12)$$

called the fitted curve.

The Numerical Approach:

We begin by writing (6) as

$$Y = F(X) = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 \tag{13}$$

Proceeding with the application of the Computational Group Average scheme, we assume that a particular point (X_1, Y_1) satisfies the curve represented by (13).

That is,

$$Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_1^3 \tag{14}$$

Subtracting (14) from (13), we have

$$Y - Y_1 = \beta_1 (X - X_1) + \beta_2 (X^2 - X_1^2) + \beta_3 (X^3 - X_1^3) \tag{15}$$

Dividing (15) through by $X - X_1$, this yields

$$\frac{Y - Y_1}{X - X_1} = \beta_1 + \beta_2 (X + X_1) + \beta_3 (X^2 + XX_1 + X_1^2) \tag{16}$$

We set $\frac{Y - Y_1}{X - X_1} = y$, $X + X_1 = x$ and $(X^2 + XX_1 + X_1^2) = z$ into (16) so that

$$y = \beta_1 + \beta_2 x + \beta_3 z \tag{17}$$

Taking the first point (1,193) from the data as the particular point that satisfies the curve and substituting into (16), then

$$\frac{Y - 193}{X - 1} = \beta_1 + \beta_2 (X + 1) + \beta_3 (X^2 + X + 1) \tag{18}$$

We divide the data into three groups since we are seeking the estimates of three unknowns β_1 , β_2 and β_3 for now. The groups are namely:

Group 1:

X	Y	z	x	y
1	193	-	-	-
2	226	7	3	33
3	240	13	4	23.5
4	244	21	5	17
Totals		41	12	73.5
Means		13.6667	4	73.5

Group 2:

X	Y	z	x	y
5	257	31	6	16
6	260	43	7	13.4
7	274	57	8	13.5
Totals		131	21	42.9
Means		43.6667	7	14.3

Group 3:

X	Y	z	x	y
8	297	73	9	14.86
9	350	91	10	19.63
10	420	111	11	25.22
Totals		275	30	59.71
Means		91.6667	10	19.9033

Putting all the means for the three groups into (17), these yield

$$24.5 = \beta_1 + 4\beta_2 + 13.6667\beta_3 \tag{19}$$

$$14.3 = \beta_1 + 7\beta_2 + 43.6667\beta_3 \tag{20}$$

And;

$$19.9033 = \beta_1 + 10\beta_2 + 91.6667\beta_3 \tag{21}$$

which constitutes a system of linear equations with three unknowns.

By putting (19) – (21) in matrix form and solving for the unknowns, using Gaussian Elimination method, we obtain

$$\beta_1 = 62.2206, \beta_2 = -12.18, \text{ and } \beta_3 = 0.8780$$

We solve for β_0 by putting the values of β_1, β_2 and β_3 into (18), so that

$$\frac{Y-193}{X-1} = 62.2206 - 12.18(X+1) + 0.8780(X^2 + X + 1) \tag{22}$$

On expanding (22) and solving for Y, then

$$Y = 0.8780X^3 - 12.18X^2 + 62.2206X + 142.096$$

That is,

$$Y = 142.096 + 62.2206X - 12.18X^2 + 0.8780X^3$$

which is identical to the estimated production cost function curve in (12)

RESULTS AND DISCUSSIONS

The OLS approach gave parameter estimates $\hat{\beta}_0 = 141.7667, \hat{\beta}_1 = 63.4776,$
 $\hat{\beta}_2 = -12.9615, \text{ and } \hat{\beta}_3 = 0.9396$ giving rise to an estimated curve
 $Y = 141.7667 + 63.4776X - 12.9615X^2 + 0.9396X^3$

The numerical approach gave unknown parameter values

$\beta_0 = 142.096, \beta_1 = 62.2206, \beta_2 = -12.18, \text{ and } \beta_3 = 0.8780$ which also led to the curve

$$Y = 142.096 + 62.2206X - 12.18X^2 + 0.8780X^3$$

By comparison, we observed that the values of the parameter estimates are very close to one another in both cases but however not exactly the same. This could be due to the errors incurred by the varying assumptions in the underlying principles of both methods.

The results *obtained* by the numerical approach also satisfy the curve. Hence, the numerical approach is equally accurate and consequently valid.

Conclusion and Recommendation:

We conclude that the computational group average (numerical) approach is valid and we recommend it as an alternative approach for estimation of polynomial regression models. We also recommend it for handling estimation in other non-linear models especially cases of small samples.

REFERENCES

- Gujarati, D.N., 2003. Basic Econometrics 4th Edition Palgrave Houndmills, ISBN: 0-07-112342-3, pp: 227-578.
- Kandasamy, P., *et al*, 1997. Numerical Methods 2nd Edition S Chand and Co.Ltd., ISBN: 81-219-1438-8, pp: 9-13 and 122-124.
- Koutsoyiannis, A., 2001. Theory of Econometrics, 2nd edition Palgrave Houndmills, ISBN: 0-333-77822-7, pp: 117-121.
- Kreuzig, E., 2004. Advanced Engineering Mathematics, 8th Edition Donford United, ISBN: 81-219-1438-8, pp: 332-339.