Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.2, No.2, 2012

Polynomial Regression Model of Making Cost Prediction In Mixed Cost Analysis

Isaac, O. Ajao (Corresponding author)
Department of Mathematics aand Statistics, The Federal Polytechnic, Ado-Ekiti,
PMB 5351, Ado-Ekiti, Ekiti state, Nigeria.
Tel: +2348035252017 E-mail: isaac_seyi@yahoo.com

Adedeji, A. Abdullahi
Department of Mathematics aand Statistics, The Federal Polytechnic, Ado-Ekiti,
PMB 5351, Ado-Ekiti, Ekiti state, Nigeria.
Tel: +2348062632084 E-mail: anzwers2003@yahoo.com

Ismail, I. Raji
Department of Mathematics aand Statistics, The Federal Polytechnic, Ado-Ekiti,
PMB 5351, Ado-Ekiti, Ekiti state, Nigeria.
Tel: +2348029023836 E-mail: rajimaths@yahoo.com

Abstract

Regression analysis is used across business fields for tasks as diverse as systematic risk estimation, production and operations management, and statistical inference. This paper presents the cubic polynomial least square regression as a robust alternative method of making cost prediction in business rather than the usual linear regression.The study reveals that polynomial regression is a better alternative with a very high coefficient of determination.

Keywords: Polynomial regression, linear regression, high-low method, cost prediction, mixed cost.

1. Introduction

Current practice in teaching regression analysis relies on the investigation of data sets for users with techniques that allow description and inference. There are many alternatives, however, for actual learner computation of regression coefficients and summary statistics. Kmenta (1971) presents a computational design that allows users to complete the calculations with only a pencil and paper. Brigham (1968) suggests that learners might simply construct a scatter plot and a ruler to visually approximate the regression line. Gujarati (2009) recommends the use of statistical packages which are now easily accessible to users on mainframe and micro computers (Mundrake, G.A., & Brown, B.J. (1989)).

Mixed costs have both a fixed portion and a variable portion. There are handful of methods used by managers to break mixed costs in the two manageable components - fixed and variable costs. The process of breaking mixed costs into fixed and variable portions allow us to use the costs to predict and plan for the future since we have a good insight on how these costs behave at various activity levels. We often call the process of separating mixed cost into fixed and variable component, cost estimation. The methods commonly used are the Scatter graph, High-low method, and the Ordinary least square linear regression. The goal of cost estimation is to determine the amount of fixed and variable costs so that a cost equation can be used to predict future costs.

2. Data and method

The high-low method uses the highest and the lowest activity levels over a period of time to estimate the portion of a mixed cost that is variable and portion that is fixed. Because it uses only the high and low activity levels to calculate the variable and fixed costs, it may be misleading if the high and low activity levels are not repreentative of the normal activity. The high-low method is most accurate when the high and low levels of activity are representation of the majority of the points.

Variable cost per unit (**b**) = $\frac{y_2 - y_1}{x_2 - x_1}$

Where y_2 = the total cost at highest level of activity

y_1 = the total cost at lowest level of activity

x_2 = are the number of units at highest level of activity; and

x_1 = are the number of units at highest level of activity

In other words, variable cost per unit is equal to the slope of the cost level line (i.e. change in total cost / change in number of units produced).

Total fixed cost (**a**) = $y_2 - bx_2 = y_1 - bx_1$

The high-low method can be quite misleading. The reason is that cost data are rarely linear and inferences are based on only two observations, either of which could be statistical anomaly or outlier. The goal of least squares is to define a line so that it fits through a set of points on a graph. Where the cummulative sum of squared distance between the points and the line is minimized, hence the name "least squares".

2.2 Polynomial Regression model

In statistics, polynomial regression is a form of linear regression in which the relationship between the independent variable *x* and the dependent variable *y* is modeled as an nth order polynomial. Polynomial regression fits a nonlinear relationship between the value of *x* and the corresponding conditional mean of *y*, denoted as $\mathrm{E}(y/x)$ (Fan, Jianqing (1996)) and (Magee, Lonnie (1998)). Although polynomial fits a non linear model to the data, as statistical estimation problem it is linear, in the sense that the regression function $\mathrm{E}(y/x)$ is linear in the unknown parameters that are estimated from the data.

2.3 The model

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^{\,2} + e_i \qquad i = 1, 2, \ldots n. \qquad \text{(i)}$$

Matematically a parabola is represented by the equation (i), also known as quadratic function, or more generally, a second-degree polynomial in the variable *x*, the highest power of of *x* represents the degree of the polynomial. If x^3 were added to the preceeding function (Gujarati, 2009) and (Studenmund, A.H., & Cassidy, H.J. (1987)), it would be a third-degree polynomial, and so on.
The stochastic version of equation (i) may be written as

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^{\,2} + \beta_3 x_i^{\,3} + e_i \qquad i = 1, 2, \ldots n \qquad \text{(ii)}$$

Which is called a second-degree polynomial regression
The general kth degree polynomial regression is written as:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^{\,2} + \ldots + \beta_k x_i^{\,k} + e_i \qquad i = 1, 2, \ldots n$$

where

β_0 , $\beta_1, \cdots \beta_\mathrm{k}$ are the parameters of the model,
ε_i is a random error term.

3. Data Presentation and Analysis

All analyses were done using MINITAB 11. The scattergram in fig(i) suggests the type of regression model that will fit the data in the table above. From this figure it is clear that the relationship between total cost and output resembles the elongated S-curve. It is noticed that the total cost curve first increases gradually and then rapidly, as predicted by the celebrated law of diminishing returns. This S-shape of the total cost curve can be captured by the following cubic or third-degree polynomial:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^{\,2} + \beta_3 x_i^{\,3} + e_i \qquad i = 1, 2, \ldots n$$

Where *y* = total cost and
x = output

3.1 Using the High-Low method

Variable cost per unit (slope) = $\dfrac{2\ 000\ 000 - 500\ 000}{175\ 000 - 60\ 000} = 13.04\, per\ unit$, that is ₦13.04 per unit

TC = FC + VC (X)

Where X = number of units

Using: Total cost (TC) = ₦ 2 000 000

Variable cost per unit (VC) = ₦ 13.04 and

X = 175 000

To obtain total fixed cost (FC)

₦2 000 000 = FC + ₦ 13.04 (175 000)

FC =₦ 2 000 000 – ₦2 282 000 = - ₦ 282 000.

The line of best fit from the above equations becomes:

$$TC = -₦\,282\,000 + ₦\,13.04\,(X) \qquad \text{(vi)}$$

The negative amount of fixed costs is not realistic and leads me to believe that either the total costs at either the high point or at the low point are not representative. The high low method of determining the fixed and variable portions of a mixed cost relies on only two sets of data: the costs at the highest level of activity, and the costs at the lowest level of activity. If either set of data is flawed, the calculation can result in an unreasonable, negative amount of fixed cost. It is possible that at the highest point of activity the costs were out of line from the normal relationship—referred to as an outlier.

4. Discussion of Results

The R-Square value is a statistical calculation that characterizes how well a particular line fits a set of data. As a general rule, the closer R^2 is to 1.00 the better; as this would represent a perfect fit where every point falls exactly on the resulting line. The models with the lowest P-value and highest R^2 which are 0.0000895 and 0.874 are the linear and polynomial cubic regression models respectively (table 4).

The negative amount of fixed costs is not realistic and leads me to believe that either the total costs at either the high point or at the low point are not representative. The high low method of determining the fixed and variable portions of a mixed cost relies on only two sets of data: the costs at the highest level of activity, and the costs at the lowest level of activity. If either set of data is flawed, the calculation can result in an unreasonable, negative amount of fixed cost. It is possible that at the highest point of activity the costs were out of line from the normal relationship—referred to as an outlier. All these are indications of it's crude and unscientific nature.

5. Conclusion and Recommendation

Based on the results of the analyses it can be concluded that Polynomial regression model is better than the conventional Linear regression and High-Low methods, especially when analysing data relating to cost and production functions.

It is obvious that Linear and Quadratic models are not too bad for prediction with respect to the data used in this research paper, but the Cubic polynomial regression is better. It is therefore recommended that data analysts should endeavour to always plot a simple scatter diagram before using any regression model in order to know the type of relationship that exists between the variable of interest.

References

Brigham, E.F. (1986). Fundamental of financial management (4th ed.). Chicago: Dryden Press.

Fan, Jianqing (1996). "1.1 From linear regression to nonlinear regression". *Local Polynomial Modelling and Its Applications*. Monographs on Statistics and Applied Probability. Chapman & Hall/CRC

Gujarati, D.N. and Porter, D.C. (2009). Basic Econometrics. New York: McGraw-Hall.

http://www.studyzone.org/testprep/math4/d/linegraph4l.cfm: Data on Monthly unit production and the associated costs

Kmenta, J. (1971). Elements of econometrics. New York: Macmillan

Magee, Lonnie (1998). "Non-local Behavior in Polynomial Regressions". *The American Statistician* (American Statistical Association) **52** (1): 20–22.

Mundrake, G.A., & Brown, B.J. (1989). Applicacation of microcomputer software to university level course instruction. Journal of Education for Business, 64(3), 124-128.

Stein, S.H. (1990). Understanding Regression Analysis. Journal of Education for Business, 65(6) 264-269.

Studenmund, A.H., & Cassidy, H.J. (1987), Using Econometric: A practical guide. Boston: Little, Brown.

Appendix

Table 1: Monthly unit production and the associated costs

(sorted from low to high)

months	Units (x)	Cost (y)
Oct	60 000	₦ 500 000
Nov	65 000	₦ 940 000
Mar	75 000	₦ 840 000
Sept	80 000	₦ 910 000

ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.2, No.2, 2012

Feb	90 000	₦ 1 100 000
Dec	95 000	₦ 1 500 000
Jan	100 000	₦ 1 250 000
Aug	115 000	₦ 1 400 000
Apr	120 000	₦ 1 400 000
Jun	130 000	₦ 1 200 000
May	140 000	₦ 1 500 000
Jul	175 000	₦2 000 000

Fig.(i): The curve of the total cost

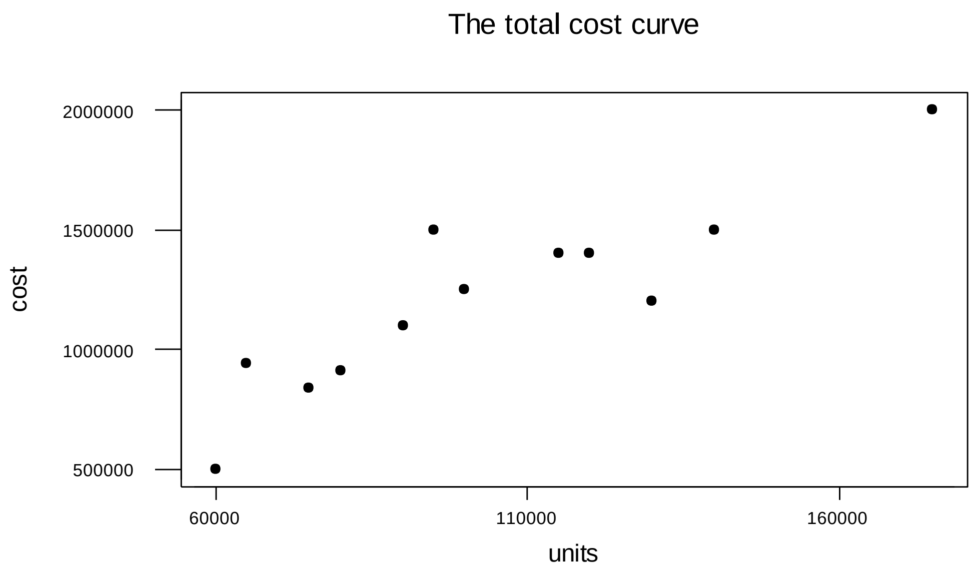


Table (2): **Regression (Linear)**

The regression equation is

y = 138533 + 10.3 x (iii)

Predictor	Coef	StDev	T	P
Constant	138533	178518	0.78	0.456
x	10.343	1.643	6.30	0.000

S = 184068 R-Sq = 79.9% R-Sq(adj) = 77.8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1.34336E+12	1.34336E+12	39.65	0.000
Error	10	3.38811E+11	33881051933		
Total	11	1.68217E+12			

Fig. (ii): Plot of the Linear regression model

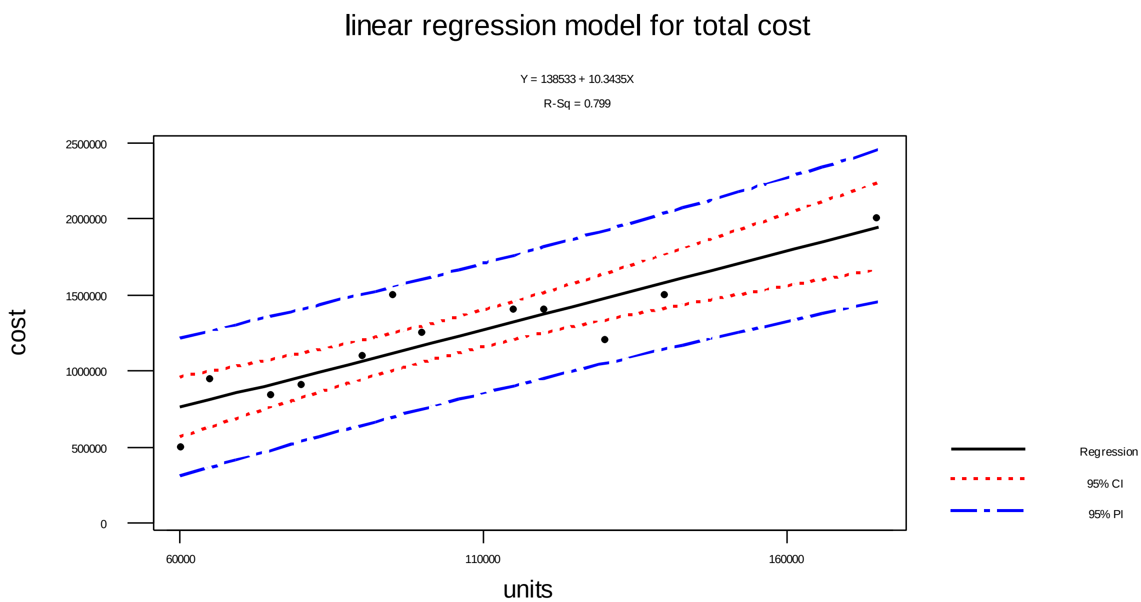


Table (3): **Polynomial Regression (Quadratic)**

Y = -136015 + 15.6406X - 2.33E-05X**2 (iv)

R-Sq = 0.804

Analysis of Variance

SOURCE	DF	SS	MS	F	P
Regression	2	1.35E+12	6.76E+11	18.4624	6.53E-04
Error	9	3.30E+11	3.66E+10		
Total	11	1.68E+12			

SOURCE	DF	Seq SS	F	P
Linear	1	1.34E+12	39.6492	8.95E-05
Quadratic	1	9.15E+09	0.249846	0.629176

Fig. (iii): Plot of the Quadratic regression model

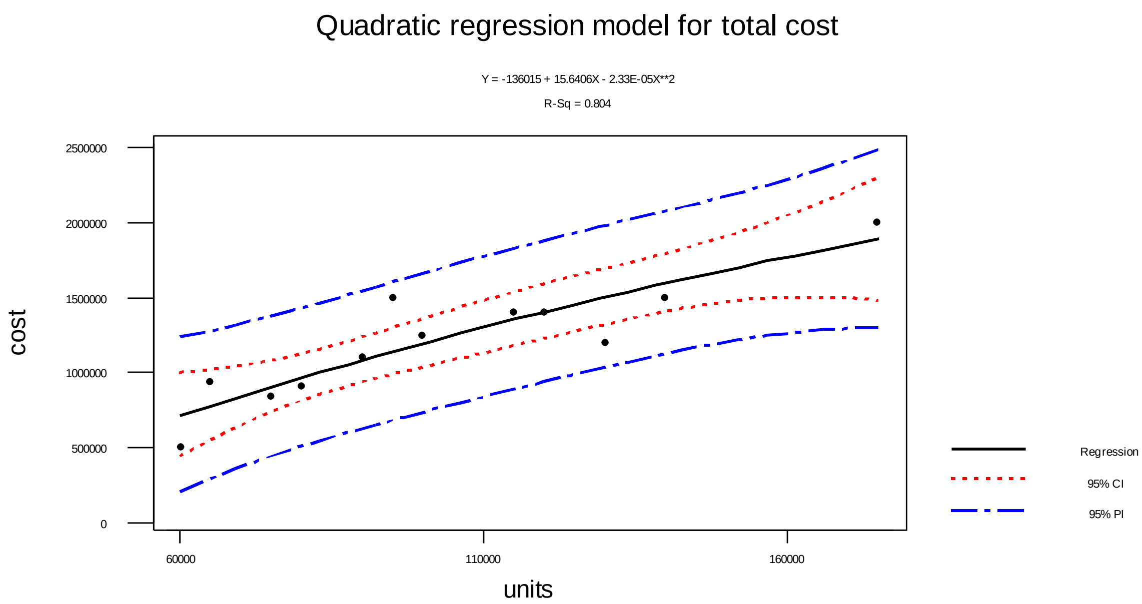


Table (4): **Polynomial Regression (Cubic)**

Y = -3888396 + 125.375X - 1.02E-03X**2 + 2.84E-09X**3 (v)

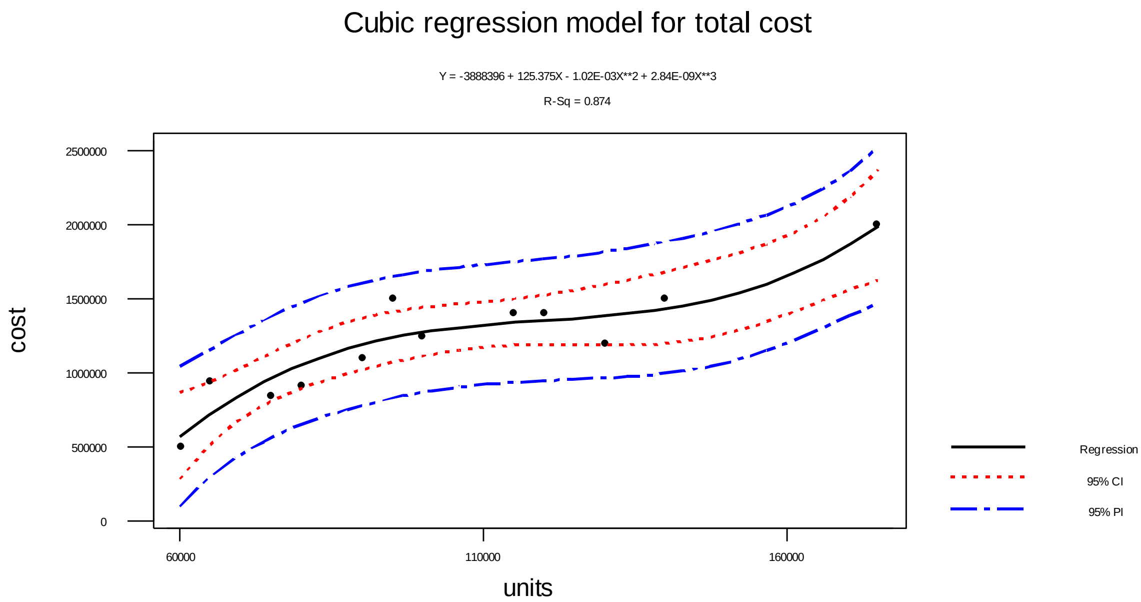
R-Sq = 0.874

Analysis of Variance

SOURCE	DF	SS	MS	F	P
Regression	3	1.47E+12	4.90E+11	18.5547	5.82E-04
Error	8	2.11E+11	2.64E+10		
Total	11	1.68E+12			

SOURCE	DF	Seq SS	F	P
Linear	1	1.34E+12	39.6492	8.95E-05
Quadratic	1	9.15E+09	0.249846	0.629176
Cubic	1	1.18E+11	4.47643	6.73E-02

Fig. (iv): Plot of the Cubic regression model


Cubic regression model for total cost
Y = -3888396 + 125.375X - 1.02E-03X**2 + 2.84E-09X**3
R-Sq = 0.874
cost
2500000
2000000
1500000
1000000
500000
0
60000
110000
160000
units
Regression
95% CI
95% PI