Chapter 12
Polynomial Regression Models

A model is said to be linear when it is linear in parameters. So the model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

and

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon$$

are also the linear model. In fact, they are the second-order polynomials in one and two variables, respectively.

The polynomial models can be used in those situations where the relationship between study and explanatory variables is curvilinear. Sometimes a nonlinear relationship in a small range of explanatory variable can also be modelled by polynomials.

Polynomial models in one variable

The k^{th} order polynomial model in one variable is given by

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + ... + \beta_k x^k + \varepsilon.$$

If $x_j = x^j,\ j = 1, 2, ..., k,$ then the model is multiple linear regressions model in k explanatory variables $x_1, x_2, ..., x_k$. So the linear regression model $y = X\beta + \varepsilon$ includes the polynomial regression model. Thus the techniques for fitting linear regression model can be used for fitting the polynomial regression model.

For example:

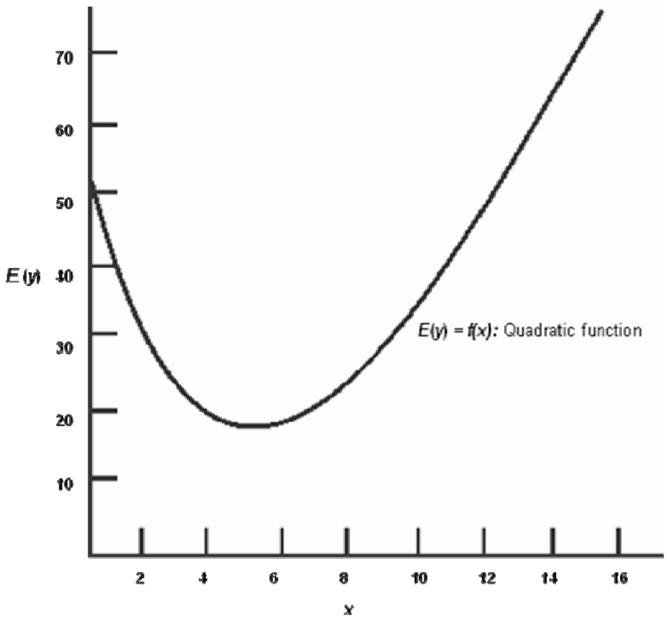
$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

or

$$E(y) = \beta_0 + \beta_1 x + \beta_2 x^2$$

is a polynomial regression model in one variable and is called a **second-order model** or **quadratic model.** The coefficients β_1 and β_2 are called the **linear effect parameter** and **quadratic effect parameter,** respectively.

The interpretation of parameter β_0 is $\beta_0 = E(y)$ when $x = 0$ and it can be included in the model provided the range of data includes $x = 0$. If $x = 0$ is not included, then β_0 has no interpretation. An example of the quadratic model is like as follows:



The polynomial models can be used to approximate a complex nonlinear relationship. The polynomial models is just the Taylor series expansion of the unknown nonlinear function in such a case.

Considerations in fitting polynomial in one variable

Some of the considerations in the fitting polynomial model are as follows:

1. Order of the model

The order of the polynomial model is kept as low as possible. Some transformations can be used to keep the model to be of the first order. If this is not satisfactory, then the second-order polynomial is tried. Arbitrary fitting of higher-order polynomials can be a serious abuse of regression analysis. A model which is consistent with the knowledge of data and its environment should be taken into account. It is always possible for a polynomial of order $(n-1)$ to pass through n points so that a polynomial of sufficiently high degree can always be found that provides a "good" fit to the data. Such models neither enhance the understanding of the unknown function nor be a good predictor.

2. Model building strategy:

A good strategy should be used to choose the order of an approximate polynomial.

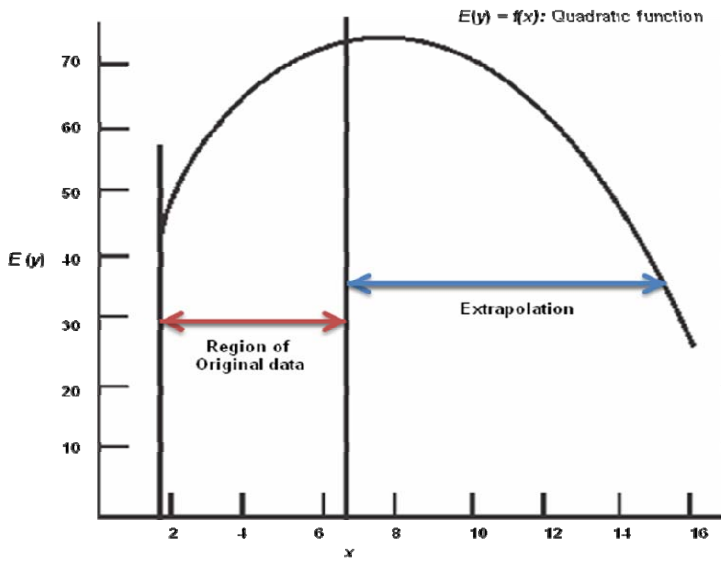
One possible approach is to successively fit the models in increasing order and test the significance of regression coefficients at each step of model fitting. Keep the order increasing until t-test for the highest order term is nonsignificant. This is called a **forward selection procedure.**

Another approach is to fit the appropriate highest order model and then delete terms one at a time starting with the highest order. This is continued until the highest order remaining term has a significant t-statistic. This is called a **backward elimination** procedure.

The forward selection and backward elimination procedures do not necessarily lead to the same model. The first and second-order polynomials are mostly used in practice.

3. Extrapolation:

One has to be very cautioned in extrapolation with polynomial models. The curvatures in the region of data and the region of extrapolation can be different. For example, in the following figure, the trend of data in the region of original data is increasing, but it is decreasing in the region of extrapolation. So predicted response would not be based on the true behaviour of the data.



In general, polynomial models may have unanticipated turns in inappropriate directions. This may provide incorrect inferences in interpolation as well as extrapolation.

4. Ill-conditioning:

A basic assumption in linear regression analysis is that rank of X-matrix is full column rank. In polynomial regression models, as the order increases, the $X'X$ matrix becomes ill-conditioned. As a result, the $(X'X)^{-1}$ may not be accurate, and parameters will be estimated with a considerable error.

If values of x lie in a narrow range, then the degree of ill-conditioning increases and multicollinearity in the columns of X matrix enters. For example, if x varies between 2 and 3, then x^2 varies between 4 and 9. This introduces strong multicollinearity between x and x^2.

5. Hierarchy:

A model is said to be hierarchical if it contains the terms x, x^2, x^3, etc. in a hierarchy. For example, the model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 + \varepsilon$$

is hierarchical as it contains all the terms up to order four. The model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_4 x^4 + \varepsilon$$

is not hierarchical as it does not contain the term of x^3.

It is expected that all polynomial models should have this property because only hierarchical models are invariant under linear transformation. This requirement is more attractive from the mathematics point of view. In many situations, the need for the model may be different. For example, the model

$$y = \beta_0 + \beta_1 x_1 + \beta_{12} x_1 x_2 + \varepsilon$$

needs a two-factor interaction which is provided by the cross-product term. A hierarchical model would need inclusion of x_2 which is not needed from the point of view of statistical significance perspective.

Orthogonal polynomials:

While fitting a linear regression model to a given set of data, we begin with a simple linear regression model. Suppose later we decide to change it to a quadratic or wish to increase the order from quadratic to a cubic model etc. In each case, we have to begin the modeling from scratch, i.e., from the simple linear regression model. It would be preferable to have a situation in which adding an extra term merely refine the model in the sense that by increasing the order, we do not need to do all the calculations from scratch. This aspect was of more importance in the pre-computer era when all the calculations were done manually. This cannot be achieved by using the powers $x^0 = 1, x, x^2, x^3 ...$ in succession. But it can be achieved by a system of orthogonal polynomials. The k^{th} orthogonal polynomial has a degree k. Such polynomials may be constructed by using Gram-Schmidt orthogonalization.

Another issue in fitting the polynomials in one variable is ill-conditioning. An assumption in usual multiple linear regression analysis is that all the independent variables are independent. In the polynomial regression model, this assumption is not satisfied. Even if the ill-conditioning is removed by centering, there may exist still high levels of multicollinearity. Such difficulty is overcome by orthogonal polynomials.

The classical cases of orthogonal polynomials of special kinds are due to Legendre, Hermite and Tehebycheff polynomials. These are **continuous orthogonal polynomials** (where the orthogonality relation involve integrating) whereas in our case, we have **discrete orthogonal polynomials** (where the orthogonality relation involves summation).

Analysis:

Consider the polynomial model of order k is one variable as

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + ... + \beta_k x_i^k + \varepsilon_i, \quad i = 1, 2, ..., n.$$

When writing this model as

$$y = X\beta + \varepsilon,$$

the columns of X will not be orthogonal. If we add another term $\beta_{k+1} x_i^{k+1}$, then the matrix $(X'X)^{-1}$ has to be recomputed and consequently, the lower order parameters $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$ will also change.

Consider the fitting of the following model:

$$y_i = \alpha_0 P_0(x_i) + \alpha_1 P_1(x_i) + \alpha_2 P_2(x_i) + ... + \alpha_k P_k(x_i) + \varepsilon_i, \; i = 1, 2, ..., n$$

where $P_u(x_i)$ is the u^{th} order orthogonal polynomial defined as

$$\sum_{i=1}^{n} P_r(x_i) P_s(x_i) = 0, \; r \neq s, r, s = 0, 1, 2, ..., k$$

$$P_0(x_i) = 1.$$

In the context of $y = X\beta + \varepsilon,$ the $X-$matrix, in this case, is given by

$$X = \begin{bmatrix} P_0(x_1) & P_1(x_1) & \cdots & P_k(x_1) \\ P_0(x_2) & P_1(x_2) & \cdots & P_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ P_0(x_n) & P_1(x_n) & \cdots & P_k(x_n) \end{bmatrix}.$$

Since this X -matrix has orthogonal columns, so $X'X$ matrix becomes

$$X'X = \begin{bmatrix} \sum_{i=1}^{n} P_0^2(x_i) & 0 & \cdots & 0 \\ 0 & \sum_{i=1}^{n} P_1^2(x_i) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sum_{i=1}^{n} P_k^2(x_i) \end{bmatrix}.$$

The ordinary least squares estimator is $\hat{\alpha} = (X'X)^{-1} X'y$ which for α_j is

$$\hat{\alpha}_j = \frac{\sum_{i=1}^{n} P_j(x_i) y_i}{\sum_{i=1}^{n} P_j^2(x_i)}, \; j = 0, 1, 2, ..., k$$

and its variance is obtained from $V(\hat{\alpha}) = \sigma^2 (X'X)^{-1}$ as

$$Var(\hat{\alpha}_j) = \frac{\sigma^2}{\sum_{i=1}^{n} \left[P_j(x_i) \right]^2}.$$

When σ^2 is unknown, it can be estimated from the analysis of variance table.

Since $P_0(x_i)$ is a polynomial of order zero, set it as $P_0(x_i)=1$ and consequently

$$\hat{\alpha}_0 = \hat{y} = \overline{y}.$$

The residual sum of squares is

$$SS_{\text{res}}(k) = SS_T - \sum_{j=1}^{k} \hat{\alpha}_j \left[\sum_{i=1}^{n} P_j(x_i) y_i \right].$$

The regression sum of squares is

$$SS_{reg}(\hat{\alpha}_j) = \hat{\alpha}_j \sum_{i=1}^{n} P_j(x_i) y_i$$

$$= \frac{\left[\sum_{i=1}^{n} P_j(x_i) y_i \right]^2}{\sum_{i=1}^{n} P_j^2(x_i)}.$$

This regression sum of squares does not depend on other parameters in the model.

The analysis of variance table, in this case, is given as follows

Source of variation	Degrees of freedom	Sum of squares	Mean squares
$\hat{\alpha}_0$	1	$SS(\hat{\alpha}_0)$	-
$\hat{\alpha}_1$	1	$SS(\hat{\alpha}_1)$	$SS(\hat{\alpha}_1)$
$\hat{\alpha}_2$	1	$SS(\hat{\alpha}_2)$	$SS(\hat{\alpha}_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\hat{\alpha}_k$	1	$SS(\hat{\alpha}_k)$	$SS(\hat{\alpha}_k)$
Residual	$n-k-1$	$SS_{\text{res}}(k)$ (by subtraction)	SS_{res}
Total	n	SS_T	

If we add another term $P_{k+1}(x_i)\alpha_{k+1}$ in the model, then the model is

$$y_i = \alpha_0 P_0(x_i) + \alpha_1 P_1(x_i) + \ldots + \alpha_{k+1} P_{k+1}(x_i) + \varepsilon_i;\ i = 1, 2, \ldots, n$$

then we just need $\hat{\alpha}_{k+1}$ and that can be obtained as

$$\hat{\alpha}_{k+1} = \frac{\sum_{i=1}^{n} P_{k+1}(x_i) y_i}{\sum_{i=1}^{n} \left[P_{k+1}(x_i) \right]^2}.$$

Notice that:

- We need not to bother for other terms in the model.
- Simply concentrate on the newly added term only.
- No re-computation of $(X'X)^{-1}$ or any other $\hat{\alpha}_j (j \neq k+1)$ is necessary due to orthogonality of polynomials.
- Thus higher-order polynomials can be fitted with ease.
- Terminate the process when a suitably fitted model is obtained.

Test of significance:

To test the significance of the highest order term, we test the null hypothesis

$$H_0 : \alpha_k = 0.$$

This hypothesis is equivalent to $H_0 : \beta_k = 0$ in polynomial regression model.

We would use

$$\begin{aligned}
F_0 &= \frac{SS_{reg}(\alpha_k)}{SS_{res}(k)/(n-k-1)} \\
&= \frac{\hat{\alpha}_k \sum_{i=1}^{n} P_k(x_i) y_i}{SS_{res}(k)/(n-k-1)} \\
&\sim F(1, n-k+1) \text{ under } H_0.
\end{aligned}$$

If the order of the model is changed to $(k+r)$, we need to compute only r new coefficients. The remaining coefficients $\hat{\alpha}_0, \hat{\alpha}_1, ..., \hat{\alpha}_k$ do not change due to the orthogonality property of polynomials. Thus the sequential fitting of the model is computationally easy.

When X_i are equally spaced, the tables of orthogonal polynomials are available, and the orthogonal polynomials can be easily constructed.

First 7 orthogonal polynomials are as follows:

Let d be the spacing between levels of x and $\{\lambda_j\}$ be the constants chosen so that polynomials will have integer values. The tables are available.

$$P_0(x_i) = 1$$

$$P_1(x_i) = \lambda_1 \left[\frac{x_i - \bar{x}}{d} \right]$$

$$P_2(x_i) = \lambda_2 \left[\left(\frac{(x_i - \bar{x})}{d} \right)^2 - \left(\frac{n^2 - 1}{12} \right) \right]$$

$$P_3(x_i) = \lambda_3 \left[\left(\frac{(x_i - \bar{x})}{d} \right)^3 - \left(\frac{x_i - \bar{x}}{d} \right) \left(\frac{3n^2 - 7}{20} \right) \right]$$

$$P_4(x_i) = \lambda_4 \left[\left(\frac{(x_i - \bar{x})}{d} \right)^4 - \left(\frac{x_i - \bar{x}}{d} \right)^2 \left(\frac{3n^2 - 13}{14} \right) + \frac{3(n^2 - 1)(n^2 - 9)}{560} \right]$$

$$P_5(x_i) = \lambda_5 \left[\left(\frac{(x_i - \bar{x})}{d} \right)^5 - \frac{5}{18}\left(n^2 - 7\right) \left(\frac{x_i - \bar{x}}{d} \right)^3 + \frac{1}{1008}\left(15n^4 - 230n^2 + 407)\right) \left(\frac{x_i - \bar{x}}{d} \right) \right]$$

$$P_6(x_i) = \lambda_6 \left[\left(\frac{(x_i - \bar{x})}{d} \right)^6 - \frac{5}{44}\left(3n^2 - 31\right) \left(\frac{x_i - \bar{x}}{d} \right)^4 + \frac{1}{176}\left(5n^4 - 110n^2 + 329)\right) \left(\frac{x_i - \bar{x}}{d} \right)^2 \right.$$
$$\left. - \frac{5}{14784}(n^2 - 1)(n^2 - 9)(n^2 - 25) \right]$$

An example of the table for $n = 5$ is as follows:

x_i	P_1	P_2	P_3	P_4
1	-2	2	-1	1
2	-1	-1	- 2	-4
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
5				
$\sum_{i=1}^{n} \left\{ P_j(x_i) \right\}^2$	10	14	10	70
λ	1	1	$\frac{5}{6}$	$\frac{35}{12}$

The orthogonal polynomials can also be constructed when x's are not equally spaced.

Piecewise polynomial (Splines):

Sometimes it is exhibited in the data that a lower order polynomial does not provide a good fit. A possible solution in such a situation is to increase the order of the polynomial, but it may always not work. The higher-order polynomial may not improve the fit significantly. Such situations can be analyzed through residuals, e.g., the residual sum of squares may not stabilize, or the residual plots fail to explain the unexplained structure. One possible reason for such happening is that the response function has different behaviour in different ranges of independent variables. This type of problems can be overcome by fitting an appropriate function in different ranges of the explanatory variable. So polynomial will be fitted into pieces. The spline function can be used for such fitting of the polynomial in pieces.

Splines and knots:

The piecewise polynomials are called splines. The joint points of such pieces are called knots. If the polynomial is of order k, then the spline is a continuous function with $(k-1)$ continuous derivatives. For this, the function values and first $(k-1)$ derivatives agree at the knots.

Cubic spline:

For example, consider a cubic spline with h knots. Suppose the knots are $t_1 < t_2 < ... < t_h$ and cubic spline has continuous first and second derivatives at these knots. This can be expressed as

$$E(y) = S(x) = \sum_{j=0}^{3} \beta_{oj} x^j + \sum_{i=1}^{h} \beta_i (x - t_i)_+^3$$

where

$$(x - t_i)_+ = \begin{cases} x - t_i & \text{if } x - t_i > 0 \\ 0 & \text{if } x - t_i \le 0. \end{cases}$$

It is assumed that the position of knots are known. Under this assumption, this model can be fitted using the usual fitting methods of regression analysis like least-squares principle.

In case, the knot positions are unknown; then they can be considered as unknown parameters which can be estimated. But in such situation, the model becomes non-linear, and methods of non-linear regression can be used.

Issue of number and position of knots:

It is not so simple to know the number and position of knots in a given set of data. It is tried to keep the number of knots as minimum as possible, and each segment should have minimum four or five data points. There should not be more than one extreme point and one point of inflexion in each segment. If such points are to be accommodated, then it is suggested to keep the extreme point in the center of the segment and point of inflexion near the knots.

It is also possible to fit the polynomials of different orders in each segment and to impose different continuity restrictions at the knots. Suppose it is to be accomplished in a cubic spline model. If all $(h+1)$ pieces of the polynomial are cubic, then a cubic spline model without continuity restrictions is

$$E(y) = S(x) = \sum_{j=0}^{3} \beta_{oj} x^j + \sum_{i=1}^{h} \sum_{j=0}^{3} \beta_i (x - t_i)_+^j$$

where

$$(x - t_i)_+^0 = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \le 0. \end{cases}$$

If the term $\beta_{ij}(x - t_i)_+^j$ is in the model, then j^{th} derivative of $S(x)$ at t_i is discontinuous.

If the term $\beta_{ij}(x - t_i)_+^j$ is not in the model, then j^{th} derivative of $S(x)$ is continuous at t_i.

So the model is fitted better when required continuity restrictions are fewer because then more parameters will be included in the model.

If more continuity restrictions are needed, then it indicates that the model is not well fitted, but the finally fitted curve will be smoother. The test of hypothesis in multiple regression model can be used to determine the order of polynomial segments and continuity restrictions.

Example:

Suppose there is only one knot at t in a cubic spline without continuity restrictions given by

$$E(y) = S(x) = \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \beta_{03}x^3 + \beta_{10}(x-t)_+^0 + \beta_{11}(x-t)_+^1 + \beta_{12}(x-t)_+^2 + \beta_{13}(x-t)_+^3 \ .$$

The term involving β_{10}, β_{11} and β_{12} are present in the model, so $S(x)$, its first derivative $S'(x)$ and second derivative $S''(x)$ are not necessarily continuous at t. Next question arises now is to judge that the quality of fit will reduce the quality of fit. This can be done by the test of hypothesis as follows:

$H_0 : \beta_{10} = 0$ tests the continuity of $S(x)$

$H_0 : \beta_{10} = \beta_{11} = 0$ tests the continuity of $S(x)$ and $S'(x)$

$H_0 : \beta_{10} = \beta_{11} = \beta_{12} = 0$ tests the continuity of $S(x), S'(x)$ and $S''(x)$.

The test $H_0 : \beta_{10} = \beta_{11} = \beta_{12} = \beta_{13} = 0$ indicates that cubic spline fits data better than a single cubic polynomial over the range of the explanatory variable x.

This approach is not satisfactory if the knots are large in number as this makes $X'X$ ill-conditioned. This problem is solved by using cubic $B-$spline which are defined as

$$B_i(x) = \sum_{j=i-4}^{i} \left[\frac{(x-t_j)_+^3}{\prod\limits_{\substack{m=i-4 \\ m \neq j}}^{i} (t_j - t_m)} \right], \; i = 1, 2, ..., h+4$$

$$E(y) = S(x) = \sum_{i=1}^{h+4} \gamma_i B_i(x)$$

where γ_i's $(i = 1, 2, ..., h+4)$ are parameters to be estimated. There are eight more knots- $t_{-3} < t_{-2} < t_{-1} < t_0$ and $t_{h+1} < t_{h+2} < t_{h+3} < t_{h+4}$. Choose $t_0 = x_{\min}$, $t_{h+1} = x_{\max}$ and other knots arbitrarily.

Polynomial models in two or more variables:

The techniques of fitting of the polynomial model in one variable can be extended to the fitting of polynomial models in two or more variables.

A second-order polynomial is more used in practice, and its model is specified by

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon .$$

This is also termed as **response surface**. The methodology of response surface methodology is used to fit such models and helps in designing an experiment. This type is generally covered in the topics in the design of experiment.