Polynomial Regression Models

Johns Hopkins Engineering

625.461 Statistical Models and Regression

Module 5 – Lecture 5B

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Polynomial Regression Models (PLM)

PLM on one variable

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

PLM on two variables

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

A Quadratic Polynomial

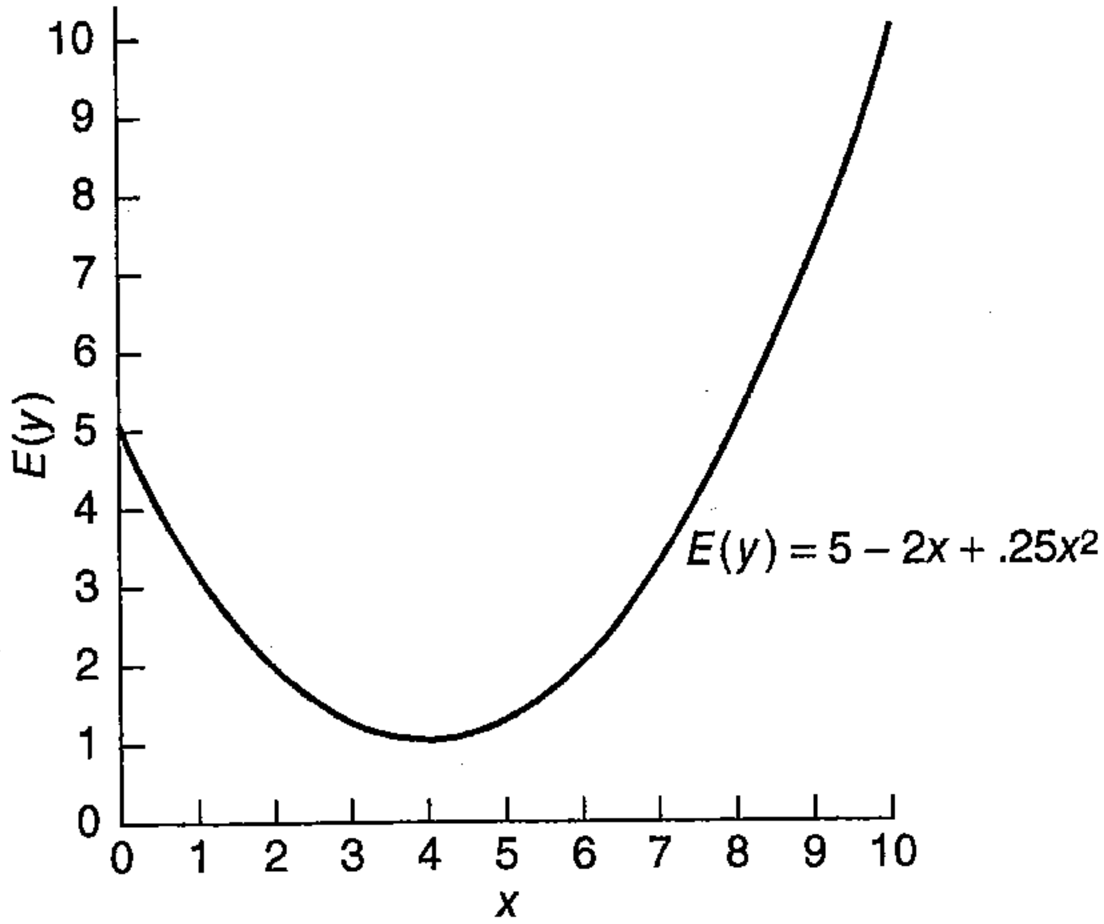


Figure 7.1 An example of a quadratic polynomial.

Important Notes with Polynomial Regression

Important to keep the order of the model as low as possible – maintain a sense of parsimony

Danger of “Extrapolation”

Danger of Extrapolation with PLM

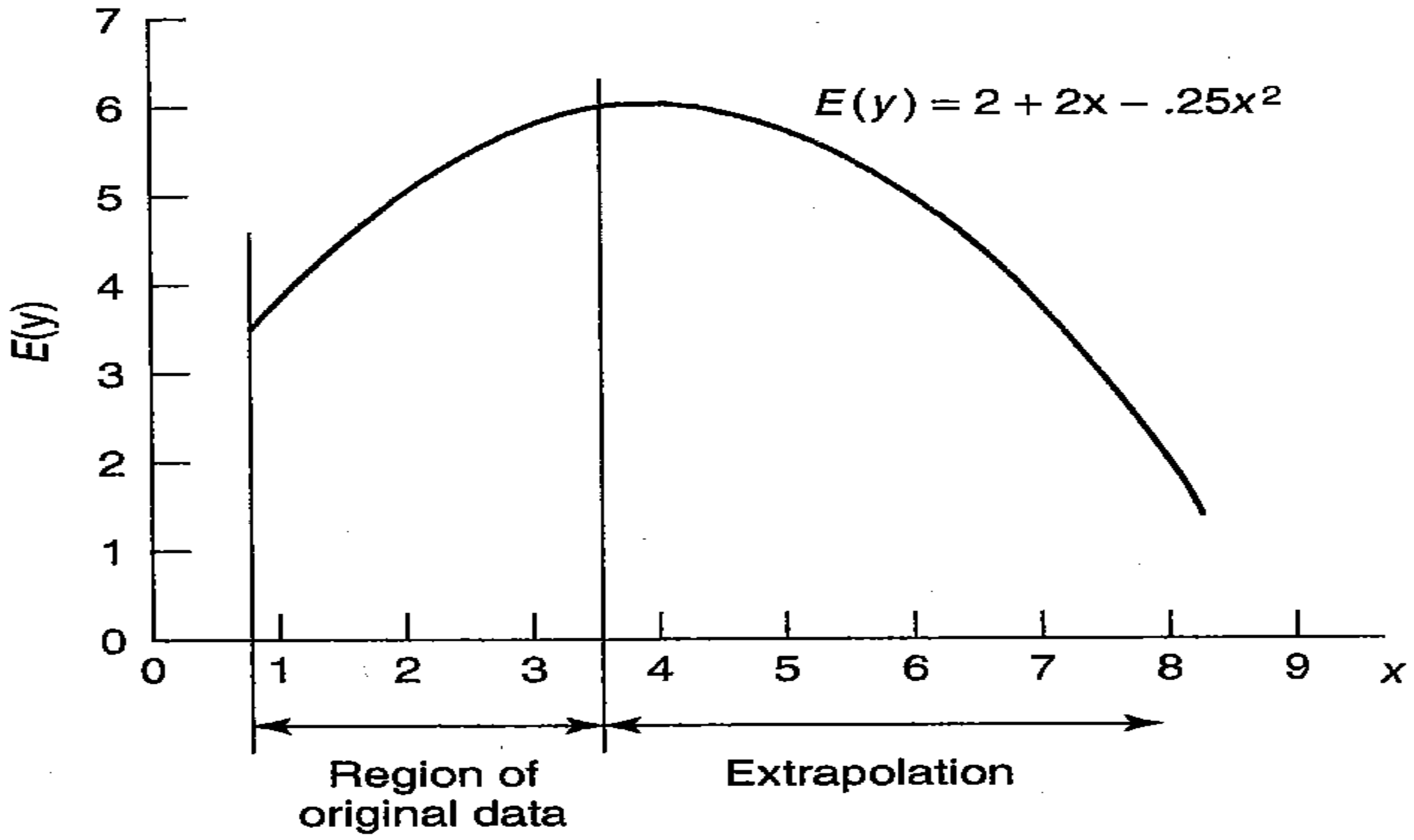


Figure 7.2 Danger of extrapolation.

Important Issues with PLM Analysis

Ill-conditioning I

As the order of the polynomial increases, the $X'X$ matrix becomes ill-conditioned. This means that the matrix inversion calculations will be inaccurate, and considerable errors may be introduced into the parameter estimates

- centering the regressor variables may help

Ill-conditioning II

If the values of x are limited to a narrow range, there can be significant ill-conditioning or multicollinearity in the columns of the X matrix. For example, if x varies between 1 and 2, x^2 varies between 1 and 4, which could create strong multicollinearity between x and x^2.

Hardwood Data (Ex 7.1, page 227 of Textbook)

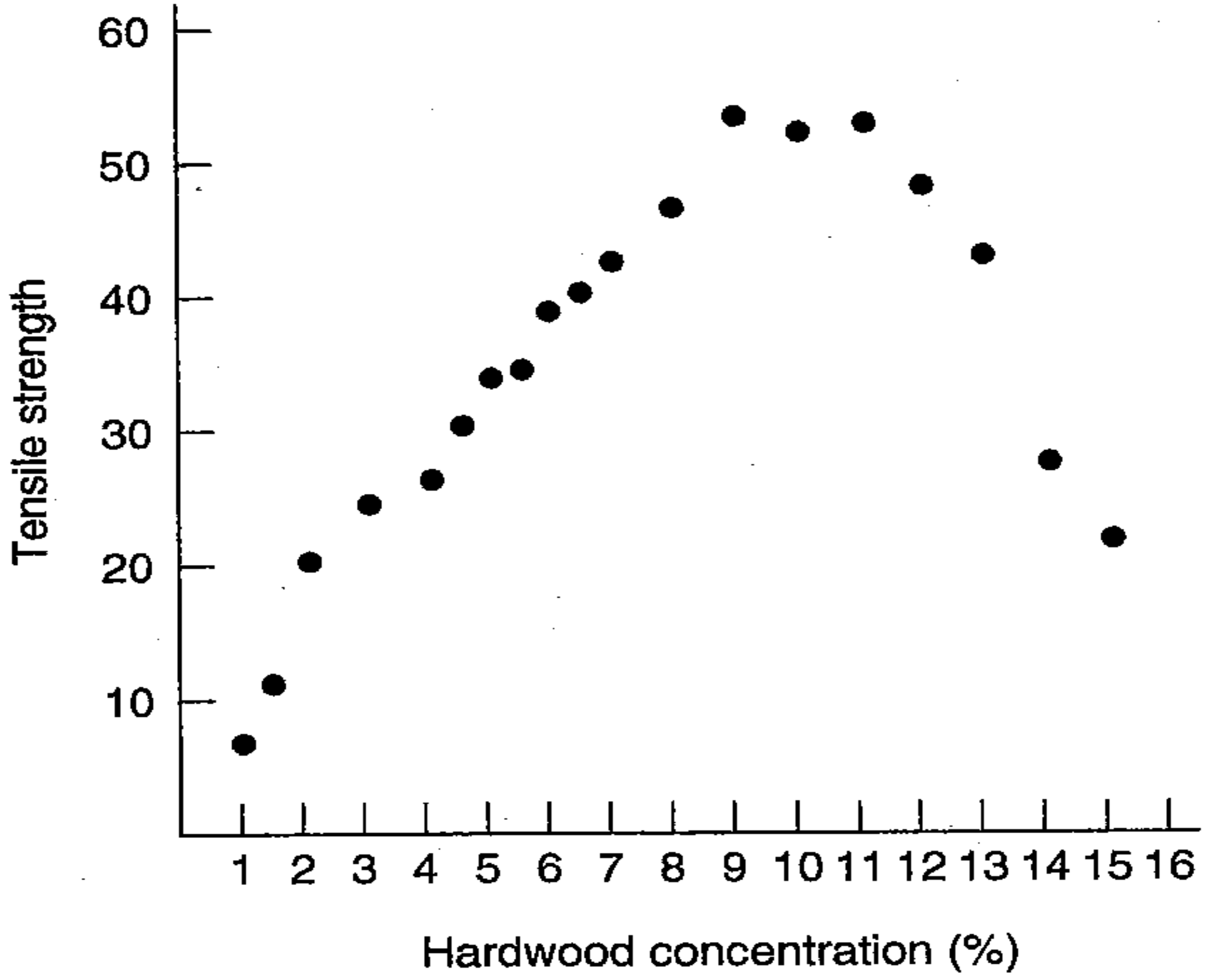
Data concerning the strength of kraft paper and the percentage of hardwood in the batch of pulp from which the paper was produced.

y: tensile strength

x: hardwood concentration

$$y = \beta_0 + \beta_1(x - \bar{x}) + \beta_2(x - \bar{x})^2 + \varepsilon$$

Hardwood Data – Scatter Diagram



Hardwood Data – Polynomial Regression Analysis

$$\hat{y} = 45.295 + 2.546(x - 7.2632) - 0.635(x - 7.2632)^2$$

TABLE 7.2 Analysis of Variance for the Qnadratic Model for Example 7.1

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P Value
Regression	3104.247	2	1552.123	79.434	4.91×10^{-9}
Residual	312.638	16	19.540		
Total	3416.885	18			

$R^2 = 0.9085$

Rejects $H_0\colon \beta_1 = \beta_2 = 0$

Hardwood Data – Polynomial Regression Analysis

We wish to investigate the contribution of the quadratic term to the model. That is, we wish to test

$$H_0: \beta_2 = 0, \quad H_1: \beta_2 \neq 0$$

If $\beta_2 = 0$, then the reduced model fit will give

$$\hat{y} = 34.184 + 1.771(x - 7.2632)$$

Under this reduced model, $MS_{\text{Res}} = 140$, $R^2 = 0.31$, $SS_{\text{R}}(\beta_1 \mid \beta_0) = 1043$.

Hardwood Data – Polynomial Regression Analysis

The extra sum of squares for testing $H_0: \beta_2 = 0$

$$SS_R(\beta_2|\beta_1, \beta_0) = SS_R(\beta_1, \beta_2|\beta_0) - SS_R(\beta_1|\beta_0)$$
$$= 3104.247 - 1043.427$$
$$= 2060.820$$

$$F_0 = \frac{SS_R(\beta_2|\beta_1, \beta_0)/1}{MS_{Res}} = \frac{2060.820/1}{19.540} = 105.47$$

$$F_{0.01,1,16} = 8.53$$

$\Rightarrow$ Conclude that $\beta_2 \neq 0$

JOHNS HOPKINS

WHITING SCHOOL *of* ENGINEERING