Nonlinear Functional Forms

Piecewise Regression

• This is another use of indicator variables in a linear model.

• Piecewise regression is used when the relationship between $Y$ and $X$ is approximated well by several different linear functions in different regions.

Pictures:

Data Example (Raw materials)
$Y =$ Unit cost (dollars) of materials
$X =$ shipment size

• Suppose there is a significant decrease in prices for shipments larger than $X_p = 500$.

• Here, $X_p$ represents a __________ __________________.

• See scatterplot for raw materials data.

A model to fit a two-piece continuous linear function:
• We see

• So when $X_1 \leq 500$, we have:

• When $X_1 > 500$, we have

• These are two linear pieces with

• Note: $\beta_2$ measures

• Note plugging $X_1 = 500$ into each equation, we get

• Fitting the regression model is done through least squares, regressing $Y$ against

Example (raw materials):

Fitted equation:

Interpretation of $b_1$ and $b_2$: 
Extensions: This approach works for 3 or more pieces. If we have changepoints at $X = 500$ and $X = 800$, the model is:

- We can fit a piecewise regression if we believe there is a discontinuity at the changepoint.

Example:

We use the model:
• Again, $\beta_2$ measures the difference in the slopes of the two pieces.

• Here, $\beta_3$ measures the

• If $\beta_3 = 0$,

(can test $H_0: \beta_3 = 0$)

**Example** (raw materials):

Fitted equation:

• If the changepoint $X_p$ is unknown, one simple approach is to fit piecewise regressions with a series (grid) of changepoint values and pick the changepoint that produces the smallest SSE (see R function).
Chapter 13: Nonlinear Regression

• Sometimes the data or underlying theory show a nonlinear relationship between \( Y \) and \( X \).

• We could try polynomial regression or using transformations of the variables, but sometimes these are also unsatisfactory. (See example scatterplot of injured patient data).

• A nonlinear regression model is of the form:

where the specified mean response function

• Sometimes a nonlinear mean response function is __________ ____________, i.e., it can be linearized by a transformation.

Example:
• If εᵢ* has “nice” characteristics (normality, constant variance), then it’s better to work with the linearized model.

• But if our model has the additive error structure:

and this εᵢ is normal with constant variance, then linearizing will ruin the “nice” error structure.

• It’s better to use nonlinear regression in that case.

• Some nonlinear models are not intrinsically linear:

Examples:

(1)

(2)

• For these models, we still assume Y is a continuous (usually normal) r.v., but the deterministic part of the relationship between Y and X is nonlinear.
Fitting the Nonlinear Model (Estimating the Parameters)

• Again, we can use least squares:

• Or assuming normal errors, we can use maximum likelihood.

Problem: It is not typically possible to analytically derive nice expressions for the regression estimates.

• We must use numerical optimization methods to either minimize the least-squares criterion or maximize the likelihood.

• These methods iteratively search across possible parameter values until the “best” estimates are found.

Search methods available in SAS:

(1)

(2)

(3)

Description of Gauss-Newton Method

• First we must choose initial estimates
• These may be selected based on previous knowledge, theoretical expectations, or a preliminary search.
• (In practice, we may use several initial guesses.)

• Use Taylor series approximation of mean response function (a Taylor series expansion around

• Then we can write the matrix “equation”:

• Estimate the unknown $\beta^{(0)}$ by least squares, obtaining $\hat{b}^{(0)}$ is the

• Then let our “revised estimates”
• Compare

• If $\text{SSE}^{(1)}$ is lower (better), then repeat the process, get

• Continue procedure until the difference in SSE: $\text{SSE}^{(s+1)} - \text{SSE}^{(s)}$, becomes negligible.

• Use “final” values

Note: The Gauss-Newton method often works well, especially with well-chosen initial values.

• Sometimes the method may take a long time to converge or may not converge at all.

• The final estimates may minimize the SSE only locally, not globally.

Other Search Methods:

• “Steepest Descent” tends to work better when the initial values are far from the final values. It iteratively determines the direction in which the regression coefficient estimates should be adjusted.

• The Marquardt method is a compromise between Gauss-Newton and Steepest Descent.

• The methods may be useful if the Gauss-Newton method runs into convergence problems.
Common Nonlinear Regression Models
(and their Characteristics)

An exponential model with 2 parameters:
\[ Y_i = \gamma_1 (1 - e^{-\gamma_2 X_i}) + \epsilon_i \]

For \( \gamma_2 > 0 \), this looks like:

- When \( X = 0 \),
- As \( X \to \infty \),
- Slope of graph when \( X = 0 \) is

Another exponential model with 2 parameters:
\[ Y_i = \gamma_1 e^{\gamma_2 X_i} + \epsilon_i \]

For \( \gamma_1 > 0, \gamma_2 < 0 \), this looks like:

- At \( X = 0 \),
- As \( X \to \infty \),
• Using another parameter could shift the function up or down: 
\[ Y_i = \gamma_0 + \gamma_1 e^{\gamma_2 X_i} + \epsilon_i \]

• The plot looks very different for \( \gamma_1 < 0 \)
(see Fig. 13.1(a), p. 512)

• Exponential models are often used in growth/decay studies.

• A Logistic Regression Model allows for an “S-shaped” curve:
\[ Y_i = \frac{\gamma_0}{1 + \gamma_1 e^{\gamma_2 X_i}} + \epsilon_i \]

For \( \gamma_0 > 0, \gamma_1 > 0, \gamma_2 < 0 \), this looks like:

• At \( X = 0 \),

• As \( X \to \infty \),
For $\gamma_2 > 0$, this logistic curve is

$$Y_i = \frac{\gamma_1 X_i}{X_i + \gamma_2} + \varepsilon_i,$$

where $\gamma_1 > 0, \gamma_2 > 0$.

- The Logistic Model is often used for population studies.

The Michaelis-Menten Model is a popular nonlinear model for enzyme kinetics to relate the initial reaction rate $Y$ to the initial substrate concentration $X$.

$$Y = \frac{\gamma_1 X}{X + \gamma_2},$$

where $\gamma_1 > 0, \gamma_2 > 0$.

- When $X = 0$,

- As $X \to \infty$,

- At $X = \gamma_2$,

- Knowledge of the meaning of the parameters allows us to use “reasonable” initial values for their estimates.
Example (Injured Patients Data):

\( Y = \text{prognosis for recovery (large is good, 0 = worst)} \)
\( X = \text{number of days in the hospital} \)

- We expect patients with longer stays in the hospital to have __________ diagnoses.

- We expect \( Y \) to be ____________ when \( X = 0 \) (no days in hospital).

- Plot of data shows

- We will use the model:

- Gauss-Newton method in SAS yields final estimates

Estimated regression function:

Inference About Parameters

- Standard methods of inference are not valid in nonlinear regression.
- But for large samples, estimators are approximately normal and approximately unbiased.

- In this case, we can use Hougaard’s statistic (which estimates the skewness of the estimators’ sampling distributions) to check their approximate normality.
Rules of thumb:

- Bootstrapping can also be useful for assessing the nature of the sampling distribution of the estimators.

Notes:

- $R^2$ in nonlinear regression is not a meaningful statistic.

- Residual plots (against fitted values), and a normal Q-Q plot of the residuals, can again be useful for diagnostics.