Alternatives to Linear Models

Bruce A Craig

Department of Statistics
Purdue University
Outline

- Linear Model Usefulness
- Alternative Approaches
- Examples
  - Generalized least squares
  - Piecewise linear model

Material mentioned in Chapter 1 of Faraway textbook
Linear Models’ Usefulness

- Focus is on relating the mean to a set of predictors
- Models generally easy to fit → no convergence issues
- Sampling dists rely more so on means being Normally distributed rather than the data themselves (CLT)
- For skewed data, variance often increases/decreases with mean so a transformation will stabilize variance and improve Normality.
- Nonlinear relationships often well-described by polynomials
- All models are wrong. Are the violations such that useful information cannot be extracted?
Alternative Approaches

What to do if model conds are not approximately met?

- Transformation of $Y$ or $X$’s
- Weighted least squares
- Bootstrapping
- Robust regression
- Regression trees
- Piecewise linear model
- Additive models
- Nonparameteric Regression
- Penalized regression
- Linear mixed model
- Generalized linear model
- Generalized linear mixed model
Case #1 - Generalized Least Squares

- Normal error regression model is reasonable except there are correlated errors and unequal variances
- Consider $Y = X\beta + \varepsilon$ where $\sigma^2(\varepsilon) = W^{-1}$ is known
- Can transform $Y$ based on $W$

$$W^{1/2}Y = W^{1/2}X\beta + W^{1/2}\varepsilon$$

$$Y_w = X_w\beta + \varepsilon_w$$

- Results in

$$E(\varepsilon_w) = 0$$
$$\sigma^2(\varepsilon_w) = I$$

- Use OLS to fit $Y_w$ to $X_w$

$$\hat{\beta} = (X'WX)^{-1}X'WY$$
Case #2 - Weighted Least Squares

- Normal error regression model is reasonable except for unequal error variances
- This is just a special case of Case #1 where \( W \) is a diagonal matrix.
- In terms of maximum likelihood

\[
Y_i \sim N(X_i \beta, \sigma_i^2) \quad (\text{\( \sigma_i \)'s known})
\]

\[
f_i = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left\{ -\frac{1}{2\sigma_i^2} (Y_i - X_i \beta)^2 \right\}
\]

- Want \( \beta \) that minimizes

\[
\sum_{i=1}^{n} \frac{1}{\sigma_i^2} (Y_i - X_i \beta)^2
\]
Weighted Least Squares

- Can be implemented in R using `lm()` with weight variable
- Optimal weights $\propto 1/var(Y_i)$
- In practice, rarely know the variances \textit{apriori}
- If they were known, define

$$W = \begin{bmatrix}
\frac{1}{\sigma_1^2} & 0 & \ldots & \ldots & 0 \\
0 & \frac{1}{\sigma_2^2} & \ddots & \vdots & \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\vdots & & \ddots & \ddots & \ddots \\
0 & \ldots & \ldots & \frac{1}{\sigma_{n-1}^2} & 0 \\
0 & \ldots & \ldots & 0 & \frac{1}{\sigma_n^2}
\end{bmatrix}$$
Determining the Weights

- Methods depend on data available
  - Use grouped data or approximately grouped data to estimate the variance
  - Compute the absolute (or squared) residual and determine a function for the standard deviation (variance).
  - Function could depend on predictors, fitted value/mean. Could be parametric (regression) or nonparametric (LOESS)
- Because weights often based on residuals, suggests iterative approach but convergence is quick
Linear Mixed Model Approach

- Often assuming variance/covariance matrix is a diagonal matrix whose values along the main diagonal (the variances) are either a
  - Linear function of $X$
  - Quadratic function of $X$
- This relationship along with the estimation of parameters can be done simultaneously using the appropriate covariance structure
- Will discuss in more detail when addressing linear mixed models
Case #3 - Piecewise Linear Model

- Lead-in to nonparametric regression
- At some point or points, the slope of the relationship changes
- When points known, can build into standard regression framework
- If points unknown, becomes a more difficult problem
  - How many change points?
  - Where is the location of each change point?
Lotsize versus Cost to Produce

- Expect cost to go down with size due to operating efficiencies
- Relationship may plateau or alter at high lot sizes
Possible Models

- Will assume there is a single changepoint
- Two possible models

Model 1: \( E(Cost) = \beta_0 + \beta_1 \text{lotsize} + \beta_2 \max(0, \text{lotsize} - C_x) \)

Model 2: \( E(Cost) = \begin{cases} 
\beta_0 + \beta_1 \text{lotsize} & \text{lotsize} \leq C_x \\
\beta_0 + \beta_1 C_x & \text{lotsize} > C_x 
\end{cases} \)

- Cycle through possible \( C_x \) to find the best changepoint
- Best refers to the model with the smallest \( s \)
- This is assuming \( \sigma \) remains constant. Reasonable?
R Code

regsigma = numeric(length=150)
regsigma1 = numeric(length=150)
for(i in 1:150){
  chngpt = 499+i
  
  cslope=0*(Lotsize<=chngpt)+(Lotsize-chngpt)*(Lotsize>chngpt)
  Lotsize1 = Lotsize*(Lotsize<=chngpt)+chngpt*(Lotsize>chngpt)

  jnk = lm(Cost~Lotsize+cslope, manufacturing)
  regsigma[i] <- summary(jnk)$sigma

  jnk1 = lm(Cost~Lotsize1, manufacturing)
  regsigma1[i] <- summary(jnk1)$sigma
}

R Code

chngpt = 499+which.min(regsigma1)

Lotsize1 = Lotsize*(Lotsize<=chngpt)+chngpt*(Lotsize>chngpt)
model1 = lm(Cost~Lotsize1, manufacturing)

plot(Lotsize,Cost,las=1)

critpts <- c(min(Lotsize)-5, chngpt, max(Lotsize)+5)

ypred = model1$coefficients[1]+critpts*model1$coefficients[2]

lines(critpts,ypred,col="red")
Lotsize versus Cost to Produce