## Chapter 11 Building the Regression Model III: Remedial Measures

In early chapters we discussed some remedial measures, such as transformations, to make the error distribution more normal, or to make the variances of the error terms more nearly equal. This chapters discusses additional remedial measures to deal with

(1) unequal error variances
(2) a high degree of multicollinearity
(3) influential observations

Nonparameteric regression methods (lowess and regression trees) are also considered.

### 11.1 Unequal error variances remedial measures --- Weighted Least Squares

The generalized multiple regression model:

$$Y_i=\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+\ldots+\beta_{p-1}X_{i,p-1}+\varepsilon_i$$

Notes:

1) $\varepsilon_i$~independent $N(0,\boldsymbol{\sigma_i^2})$ (This is new!)
2) $\beta_0, \beta_1, \ldots,\beta_{p-1}$ are parameters
3) $X_{i1},\ldots,X_{i,p-1}$ are known constants. The second subscript on $X_{ij}$ denotes the $j^{th}$ independent variable.
4) $i=1,\ldots n$, represents the ith trial

**Error variances known**

Define the reciprocal of the variance as the weight $w_i$: $w_i=1/\boldsymbol{\sigma_i^2}$

Let the matrix W be a diagonal matrix containing the weight $w_i$, then the weighted least squares and the maximum likelihood estimators of the regression coefficients are:

$$\mathbf{b}_w=\mathbf{(X'WX)^{-1}X'WY}$$

**Error variances unknown**

The problem becomes how to estimate $\boldsymbol{\sigma_i^2}$**.**

Estimation of variance function or standard deviation function

Since $\sigma_i^2=E\{\varepsilon_i^2\}-(E\{\varepsilon_i\})^2=E\{\varepsilon_i^2\}$, the squared residual $e_i^2$ is an estimator of $\sigma_i^2$. Furthermore, the absolute residual $|e_i|$ is an estimator of the standard deviation $\sigma_i$.

Weighted Least Squares Estimation Process

(1) Fit the regression model by unweighted least squares and analyze the residuals.

(2) Estimate the variance function or the standard deviation function by regressing either the squared residuals or the absolute residuals on the appropriate predictor(s).

(3) Use the fitted values from the estimated variance or standard deviation function to obtain the weights $w_i$.

$$w_i = \frac{1}{(\hat{s}_i)^2} \text{ where } \hat{s}_i \text{ is fitted value from standard deviation function}$$

$$w_i = \frac{1}{\hat{v}_i} \text{ where } \hat{v}_i \text{ is fitted value from variance function}$$

(4) Estimate the regression coefficients using these weights.

Notes: Usually iteratively reweighted least squares are recommended. That is, to iterate the weighted least squares process by using the residuals from the weighted least squares fit to reestimate the variance or standard deviation function and then obtain revised weights. Often one or two iterations are sufficient to stabilize the estimated regression coefficients.

Example on page 427.

Softwares:

R: lm(Y~X, weights=**variable**)

SAS: proc reg; model Y=X; weight **variable**; run;

**11.2 Multicollinearity remedial measures --- Ridge Regression**

Ridge regression has been proposed to remedy multicollinearity problems by modifying the model of least squares to allow biased estimators of the regression coefficients. See Figure 11.2 on page 434.

When all variables are transformed by correlation transformation, the least square normal equations are:

$$\mathbf{r}_{XX}\mathbf{b}=\mathbf{r}_{YX}$$

The ridget standardized regression estimators are obtained by introducing a biasing constant c, in the following form:

$$(\mathbf{r}_{XX}+c\mathbf{I})b^R = \mathbf{r}_{YX}$$

Then the ridge standardized regression coefficients are:

$$b^R = (\mathbf{r}_{XX}+c\mathbf{I})^{-1}\,\mathbf{r}_{YX}$$

**Choices of biasing constant c**

A commonly used method of determining c is based on the **ridge trace** and the variance inflation factors (VIF).

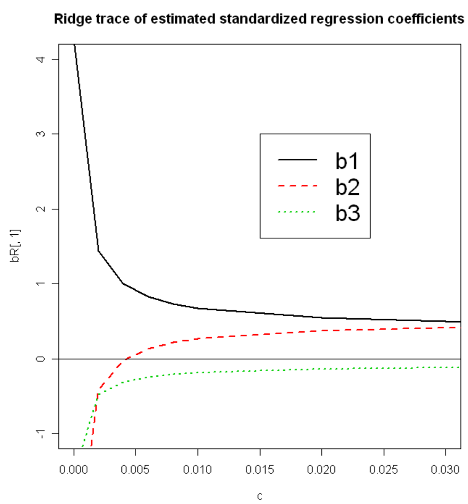
The ridge trace is a simultaneous plot of the values of the p-1 estimated ridge standardized regression coefficients for different values of c, usually between 0 and 1.

The estimated ridge coefficients may fluctuate widely as c is changed slightly from 0, and some may even change signs. Gradually these wide fluctuations cease and the magnitudes of the regression coefficients ten to move slowly toward zero as c increased further. At the same time, the values of VIF tend to fall rapidly as c is changed from 0 and gradually the VIF values also tend to change only moderately as c is increased further. One therefore examines the ridge trace and the VIF values and **chooses the smallest value of c where it is deemed that the regression coefficients first become stable in the ridge trace and the VIF values have become sufficiently small**.

**Example: Body fat example (**Page 434). Three predictor variables.

**Check on R codes:**

| c | $b_1^R$ | $B_2^R$ | $B_3^R$ | $VIF_1$ | $VIF_2$ | $VIF_3$ |
|---|---|---|---|---|---|---|
| 0 | 4.264 | -2.929 | -1.561 | 708.843 | 564.343 | 104.606 |
| 0.002 | 1.441 | -0.411 | -0.481 | 50.559 | 40.448 | 8.280 |
| 0.004 | 1.006 | -0.025 | -0.315 | 16.982 | 13.725 | 3.363 |
| 0.006 | 0.830 | 0.131 | -0.247 | 8.503 | 6.976 | 2.119 |
| 0.008 | 0.734 | 0.216 | -0.210 | 5.147 | 4.305 | 1.624 |
| 0.01 | 0.674 | 0.268 | -0.187 | 3.486 | 2.981 | 1.377 |
| 0.02 | 0.546 | 0.377 | -0.137 | 1.103 | 1.081 | 1.011 |
| 0.03 | 0.500 | 0.413 | -0.118 | 0.626 | 0.697 | 0.923 |
| 0.04 | 0.476 | 0.430 | -0.108 | 0.453 | 0.555 | 0.881 |
| 0.05 | 0.460 | 0.439 | -0.101 | 0.370 | 0.486 | 0.853 |
| 0.1 | 0.423 | 0.449 | -0.081 | 0.252 | 0.373 | 0.761 |
| 0.5 | 0.338 | 0.379 | -0.030 | 0.154 | 0.214 | 0.403 |
| 1 | 0.280 | 0.310 | -0.006 | 0.107 | 0.136 | 0.227 |




The resulting fitted model is chosen for c=0.02. Then use (7.53) to transform back to the original variables.

```
> print(c(b0,b1,b2, b3))
[1] -7.4034254  0.5553531  0.3681444 -0.1916269
```

**11.3 Remedial measures for influential cases --- Robust Regression**

(1) LAR or LAD regression: least absolute residuals (LAR) or least absolute deviation (LAD) regression, also called L1-norm regression, estimates the regression coefficients by minimizing the sum of the absolute deviations of the Y observations from their means.
(2) IRLS robust regression: iteratively reweighted least squares (IRLS) robust regression uses the weighted least squares procedure where the weights are based on how far outlying a case is (measured by the residual) and the weights are revised with each iteration.
(3) LMS regression: least median squares (LMS) replaces the sum of squared deviations in ordinary least squares by the median of the squared deviations.

**11.4 Nonparameteric regression: Lowess Method and Regression Trees**

(1) No analytic expression for the response surface is provided
(2) With more than two predictor variables, it is not possible to show the fitted response surface graphically

**LOWESS**: locally weighted regression and smoothing method. For any combination of X levels, the Lowess method fits either a first-order model or a second-order model based on the cases in the neighborhood, with more distant cases in the neighborhood receiving smaller weights. The neighborhood about a given point is defined in terms of the proportion q of cases that are nearest to the point.

**R commands**: lowess can be used when there is only one predictor variable; loess can be used when you have one to four predictor variables.
Example:

```
 ######### loess and lowess
colnames(Data) <- c("X1", "X2", "X3", "Y")
plot(Y~X1, data=Data)
lowess.fit <- lowess(Data$Y~Data$X1)
lines(lowess.fit)

loess.fit <- loess(Data$Y~Data$X1+Data$X2)
predict(loess.fit)
```

**Regression Trees:** For the case of single predictor, the range of the predictor is partitioned into segments and within each segment the estimated regression fit is given by the mean of the responses in the segment. For two or more predictors, the X space is partitioned into rectangular regions, and again, the estimated regression surface is given by the mean of the responses in each rectangle. (see the graph on page 6)

**11.5 Remedial measures for evaluating precision in nonstandard situations --- Bootstrapping**

**General procedure**

Suppose we want to evaluate the precision of an estimated regression coefficient which is obtained by fitting a regression model by some procedure. The explanation applies identically to any other estimate.

(1) Select a random sample of size n with replacement from the observed sample;
   a. Fixed X sampling: The residuals $e_i$ from the original fitting are regarded as the sample data to be sampled with replacement. After a bootstrap sample of the residuals has been obtained, the bootstrap sample residuals are added to the fitted values from the original fitting to obtain new bootstrap Y values: $Y_i^* = \hat{Y}_i + e_i^*$. The bootstrap Y* values are then regressed on the original X variables by the same procedure used initially to obtain the bootstrap estimate.
   b. Random X sampling: The pairs of X and Y data in the original sample are considered to be the data to sampled with replacement. It samples cases with replacement n times yielding a bootstrap sample of n pairs of (X*,Y*) values.

(2) Calculate the estimated regression coefficient from the bootstrap sample using the same fitting procedure as employed for the original fitting;
(3) Repeat steps (1) and (2) for a large number of times;
(4) The estimated standard deviation of all o the bootstrap estimates is an estimate of the variability of the sampling distribution of the original regression coefficient.

**Bootstrap confidence intervals**

A relatively simple procedure for setting up a 1-$\alpha$ confidence interval is the **reflection method**. The reflection method confidence interval for $\beta_1$ is based on the $(\alpha/2)100$ and $(1-\alpha/2)100$ percentiles of the bootstrap distribution $b_1^*$. These percentiles are denoted by $b_1^*(\alpha/2)$ and $b_1^*(1-\alpha/2)$, respectively. The distances of these percentiles from $b_1$, the estimate of $\beta_1$ from the original sample, are denoted by $d_1$ and $d_2$:

$$d_1 = b_1 - b_1^*(\alpha/2)$$
$$d_2 = b_1^*(1-\alpha/2) - b_1$$

The approximate 1-$\alpha$ confidence interval for b1 then is

$$b_1 - d_2 \le \beta_1 \le b_1 + d_1.$$

**Toluca Company Example** (Example on Page 460). Also see Chapter 2 Lecture Notes on page 14.