Some Remedial Measures:

Problems	Solution
1. Unequal error variance	Weighted Least Squares
2. Multicollinearity	Ridge regression
3. Influential case	Robust regression
4. Unknown response	Non-parametric regression

In order to evaluate precision in non-standard situations we often use computer intensive techniques like bootstrapping.

Unequal error Variance

Situation: Appropriate regression function has been found. However the error variance is unequal.
Possible Solution: Transform Y or the X's
Problem: The estimated relationship may change.
We may not have a linear relationship.

In these situations use WEIGHTED LEAST SQUARE techniques.

Model: $y = \beta 0+\beta 1x1 +\ldots+\beta p\text{-}1xp\text{-}1 + \varepsilon$

Where $\beta 0, ..., \beta p\text{-}1$ represent parameters
$x1, \ldots xp\text{-}1$ represent known constants
εi are independent and $\varepsilon i \sim N(0, \sigma_i^2)$

Note we are no longer assuming equal variances.

For this model above if ordinary least squares are found they will be:

- Unbiased
- Consistent
- BUT NOT Minimum Variance

If the error variances are *different* the Yi's are of different reliability.
Small variances – more reliable
Large variances – less reliable

The importance of the Yi's are inversely proportional to their variances. And we need to minimize:

Question is:

How do we estimate wi's = $1/\sigma_i^2$

Two methods:

- Estimation of the variance function or the standard deviation function
- Use of replicates or near replicates

Estimation of the Variance Function or standard deviation function:

IDEA: the magnitude of σ_i^2 or σ_i often varies in a regular fashion with one or several predictors.

Hence to find the Variance function (standard deviation function), relate σ_i^2 (σ_i) to the relevant predictors x1, x2, …, xp-1.

To do so we want to regress e_i^2 or |ei| on the predicts x1, …, xp-1. Use residual plots to determine which of the variables can enter into the model.

Procedure:

1. Fit the regression model using unweighted least squares
2. Analyze the residuals
3. Estimate the variance (standard deviation) function by regressing ei2 (|ei|) on the relevant predictors.
4. Use the predicted values (from the above regression) to estimate wi
5. Re-estimate the regression coefficients using weights.

If the WLS estimates are very different from the ordinary least squares, one may want to repeat this procedure a few times to reach stability.

(ITERARATIVELY REWIGHTED LEAST SQUARES)

Use of near replicates or replicates:

If at a particular X value replicates of y are available, one can calculate the weights directly. Using the replicates one can calculate the variance for each x level. We may also regress the sample variances on the x's and calculate the weights.

If replicates are not available "near replicates" may be used.

Inference procedures when weights are estimated:

- Variances of the estimators are obtained using the estimated weights.
- Testing confidence intervals etc can be done using s(bwk) in the usual way.
- Approximations are quite good if the sample sizes are large

Summary and Words of Caution:

- Weighted least squares are often used to correct HETEROSCADASTICITY
- This is a problem when the response function generally follows a distribution where the variance is functionally related to the mean.
- If there are major differences in the variances of the error terms the estimation of weights may be important
- R^2 is not quite so easy to interpret
- Special case of generalized least squares, when the error terms may also be correlated.