Differing Error Variances

Method A: Transform $Y$ (possibly $X$) e.g.

$$\ln(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

This creates a nonlinear relationship as it alters the variability.

Downside:

Perhaps a linear relationship is adequate.
Response $\ln(Y)$ is more difficult to interpret.

Variability increases with $X$, but a linear relationship is reasonable.
Weighted Least Squares

Choose $\beta$-vector, $b_w$, which minimizes

$$\sum_{i=1}^{n} w_i \left( Y_i - (\beta_0 + \beta_1 x_{i1} + \ldots + \beta_{p-1} x_{ip-1}) \right)^2$$

with weights $w_i \geq 0 \quad i=1, \ldots, n$.

It is more important to get close to the observed $Y_i$ with the predicted $\hat{Y}_i$ for some points (larger weights) than others.
i.e. the importance of a point is inversely proportional to the variance there.

If we have multiple data points at each fixed $X$-value. Use those values to estimate $\sigma_i^2$ with the sample variance of the $Y_i$'s at each particular $X$-value ($X$ combination).
In many situations replicates are not available. We could use a "window" of $X$-combinations close to $X_i$, and compute an estimate of the $Y$-variability among $X$-combinations that are close. But this requires a measure of closeness (in higher dimensions $(p-1)$ this is not clearly described.

\[ e_i \]

Example:
Magnitude of the residuals increases as the first predictor $X_1$ increases.
We could model the variability with

$$\sigma_i = \alpha_0 + \alpha_1 X_{i1} + \epsilon_i$$

That is, we construct a regression equation with

$|e_i|$ as response and $X_{i1}$ as predictor
analyzing the pairs $(X_{i1}, |e_i|)$ $i = 1, \ldots, n$.

Estimating the relationship

$$|e| = a_0 + a_1 X$$
Approximate C.I. for $\beta_k$ is:

$$b_k \pm t \sqrt{S^2(b_k)}$$

and test statistic for $H_0: \beta_k = 0$ vs $H_a: \not= H_0$ is:

$$\frac{b_k - 0}{\sqrt{S^2(b_k)}}.$$
Estimate $E[Y_h]$ at $\hat{X}_h$ with

\[ \hat{Y}_h = b_w' \hat{X}_h \]  

point est.

---

Remedial Measure for Multicollinearity —

Ridge Regression.

Suppose

(a) You can not add data points to uncorrelate the predictors and

(b) You want to continue to use all of the predictors.
Transform the data:

\[ X_{i,j}^* = \frac{X_{i,j} - \bar{X}_{j}}{\sqrt{n-1} S_X} \quad Y_i^* = \frac{Y_i - \bar{Y}}{\sqrt{n-1} S_Y} \]

\[ i = 1, \ldots, n \]
\[ j = 1, \ldots, p-1 \]

Now

\[
\begin{pmatrix}
X^* \times X^ *
\end{pmatrix} = X_{XX} = \text{Correlation Matrix}
\]

\[
= \begin{pmatrix}
1 & r_{12} & r_{13} & \cdots & r_{1p-1} \\
r_{12} & 1 & r_{23} & \cdots & r_{2p-1} \\
r_{13} & r_{23} & 1 & \cdots & r_{3p-1} \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
r_{1p-1} & r_{2p-1} & r_{3p-1} & \cdots & 1
\end{pmatrix}
\]

Symmetric

\((p-1) \times (p-1)\)

\[
\begin{pmatrix}
X^* \times Y^ *
\end{pmatrix} = \begin{pmatrix}
r_{YX_1} \\
\vdots \\
r_{YX_{p-1}}
\end{pmatrix} = Y_{XX} \quad (p-1) \times 1
If the $X_j$ predictors are highly correlated, then $R_{xx}$ may still be nearly singular. This produces:

(a) instability in $b^*$'s computation

(b) large estimates of $\text{Variance}(b_{xk})$

So before we solve for $b^*$, consider making $R_{xx}$ less singular! Find

$$I \approx (p-1) \times (p-1) \text{ identity matrix}$$

and $c > 0$. 
The challenge in using this method is

As $c$ increases from 0,

(i) $b_k^R$ may change rapidly at first, then plateau, eventually moving slowly toward zero.

(ii) The $VIF_k(c)$ will decrease rapidly, eventually going to one.
Here $\text{VIF}_k(c)$ is the $k, k^{th}$ element.

The choice of $c$ is subjective.

**Ridge Trace** — a simultaneous graph of the $b_k$'s $k=1, \ldots, p-1$ as a function of $0 \leq c \leq 1$.

**Example:**
2. Weighted Least Squares is a good option for leaving the data set intact and yet diminishing the influence of an extreme point (or points).

(a). Fit the usual (unweighted) least squares fit. Let $e_i$ denote the residual of the $i^{th}$ data point based on this fit.
(b) Form: \[ e_i^* = \frac{e_i}{\text{MAD}(e)} \]

(c) Find the weighted L.S. estimate \( b^w \) of \( \beta \) by minimizing

\[ \sum_{i=1}^{n} w_i \left( Y_i - (\beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}) \right)^2 \]

Find the residuals \( e_i \) from this fit.

(d) Repeat steps (b) and (c) until the estimate \( b^w \) stops changing.
The weights in the final step will indicate which data points are extreme, i.e. have been downweighted.

**Example in SAS**

<table>
<thead>
<tr>
<th></th>
<th>Iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>Price</td>
<td></td>
</tr>
<tr>
<td>Me</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td></td>
</tr>
<tr>
<td>Abs, Dev.</td>
<td></td>
</tr>
</tbody>
</table>

Downweighted Points on Iteration #3

<table>
<thead>
<tr>
<th>Point #</th>
<th>Weight</th>
</tr>
</thead>
</table>