General Setup for Nonlinear Models

Data are pairs $(Y_j, x_j), j = 1, 2, \ldots, n$.

A regression model is a specification of the distribution of $Y|x$.

Most often, we focus on the first two moments,

$E(Y_j|x_j) = f(x_j, \beta)$,

$\text{var}(Y_j|x_j) = \sigma_j^2$.

Here $x$ is $(r \times 1)$ and $\beta$ is $(p \times 1)$, with possibly $p \neq r$. 
Inferential Approaches

Mostly, scientific focus is on $\beta$.

Objective is to estimate, and test hypotheses about, $\beta$.

The variances $\sigma_j^2$ may be unequal, and will need to be estimated or modeled, to assure:

- reasonably efficient estimates of $\beta$;
- valid standard errors for and tests about $\beta$. 
Approach 1: Ordinary Least Squares (OLS)

Choose $\beta$ to minimize

$$S(\beta) = \sum_{j=1}^{n} \{ Y_j - f(x_j, \beta) \}^2 .$$

Motivation:

- “sensible” way to find good values of $\beta$.
- If we assume normality with common variance $\sigma^2$, i.e. $Y_j \sim N\{ f(x_j, \beta) , \sigma^2 \}$, equivalent to maximizing the log-lik

$$\log L = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^{n} \{ Y_j - f(x_j, \beta) \}^2 .$$
Remarks

This Gaussian likelihood is based on the conditional distribution of \( Y \mid x \). The full likelihood has a further factor, the marginal distribution of \( x_1, x_2, \ldots, x_n \). If that marginal distribution involved \( \beta \), we would want to use the full likelihood.

Minimizing \( S(\beta) \) leads to the estimating equation

\[
\sum_{j=1}^{n} \left\{ Y_j - f(x_j, \beta) \right\} f_\beta(x_j, \beta) = 0 \quad (p \times 1)
\]

where

\[
f_\beta(x, \beta) = \left\{ \frac{\partial f(x, \beta)}{\partial \beta_1}, \frac{\partial f(x, \beta)}{\partial \beta_2}, \ldots, \frac{\partial f(x, \beta)}{\partial \beta_p} \right\}^T.
\]
Approach 2: Weighted Least Squares (WLS)

If we accept that the variances are unequal, but assume that we know them up to a constant of proportionality:

\[ \sigma_j^2 = \frac{\sigma^2}{w_j}, \]

with \( \sigma^2 \) unknown, but \( w_1, w_2, \ldots, w_n \) known (unlikely!);

Then we would minimize the weighted sum of squares

\[ S_w(\beta) = \sum_{j=1}^{n} w_j \{ Y_j - f(x_j, \beta) \}^2. \]
The estimating equation becomes

\[
\sum_{j=1}^{n} w_j \{ Y_j - f(x_j, \beta) \} f_\beta(x_j, \beta) = 0
\]

Again, we find the same estimating equation by adopting the additional assumption of Gaussian distributions, and maximizing the log likelihood

\[
\log L = - \frac{1}{2} \sum_{j=1}^{n} \log \left( 2\pi \frac{\sigma^2}{w_j} \right) - \frac{1}{2\sigma^2} \sum_{j=1}^{n} w_j \{ Y_j - f(x_j, \beta) \}^2 .
\]
Remarks

Note that, for both OLS and WLS, the estimating equation

\[ \sum_{j=1}^{n} w_j \{ Y_j - f(x_j, \beta) \} f_\beta(x_j, \beta) = 0 \]

is:

- linear in the \( Y_s \), which is convenient in developing the sampling theory of the estimators;
- in general, nonlinear in the \( \beta_s \), which is inconvenient in finding numerical solutions.

MLEs may be found by direct numerical optimization of the sum of squares, but solving the estimating equations is preferred.
Approach 3: Plug-in estimates of weights

Suppose we have replicates: \( Y_{j,k} = k^{th} \) observation at \( x_j \), \( k = 1, 2, \ldots, r_j \).

If we assume that \( \text{var}(Y_{j,k}) = \sigma_j^2 \), we can estimate it unbiasedly by

\[
s_j^2 = \frac{1}{r_j - 1} \sum_{k=1}^{r_j} (Y_{j,k} - \bar{Y}_j)^2.
\]

We could then plug these into \( S_w \), and minimize

\[
S_{s^2}(\beta) = \sum_{j=1}^{n} \sum_{k=1}^{r_j} \frac{(Y_{j,k} - f(x_j, \beta))^2}{s_j^2}.
\]
Remarks

This can be rewritten

\[ S_{s^2}(\beta) = \sum_{j=1}^{n} \left\{ \bar{Y}_j - f(x_j, \beta) \right\}^2 \frac{s_j^2}{r_j} + \sum_{j=1}^{n} (r_j - 1), \]

depending only on the summary statistics \((\bar{Y}_j, s_j^2), j = 1, 2, \ldots, n\).

The estimating equation is

\[ \sum_{j=1}^{n} \frac{\{ \bar{Y}_j - f(x_j, \beta) \} f_{\beta}(x_j, \beta)}{s_j^2/r_j} = 0, \]

linear in \(\bar{Y}_j\) (but not linear in \(Y_{j,k}\)).
If $r_j$ is small, $s_j^2$ is a poor estimator of $\sigma_j^2$, so this may not work well.

We have not exploited any structure in the variation of $\sigma_j^2$. 
Approach 4: Likelihood with replicates

Again, suppose we have replicates $Y_{j,k}$ at $x_j$, $k = 1, 2, \ldots, r_j$.

As before, assuming Gaussian distributions leads to a likelihood function

$$
\log L = -\frac{1}{2} \sum_{j=1}^{n} r_j \log (2\pi \sigma_j^2) - \sum_{j=1}^{n} \sum_{k=1}^{r_j} \left\{ \frac{Y_{j,k} - f(x_j, \beta)}{2\sigma_j^2} \right\}^2.
$$

Differentiation wrt $\beta$ and $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ yields equations for all unknowns.
Remarks

Again, the likelihood can be rewritten to depend only on the summary statistics \((\bar{Y}_j, s_j^2)\), \(j = 1, 2, \ldots, n\).

The estimating equations are

\[
\sum_{j=1}^{n} \frac{\{\bar{Y}_j - f(x_j, \beta)\}}{\sigma_j^2/r_j} f_\beta(x_j, \beta) = 0,
\]

\[
\frac{1}{r_j} \sum_{k=1}^{r_j} \{Y_{j,k} - f(x_j, \beta)\}^2 = \sigma_j^2, \quad j = 1, 2, \ldots, n.
\]

These could be solved iteratively, given say initial estimates \(s_j^2\) (or 1) of \(\sigma_j^2\).
Comments on Approaches 1–4

Approach 1 (OLS) is inefficient when variances are truly unequal.

Approach 2 (WLS) requires unlikely information about variance ratios.

Approaches 3 and 4 require replication, which may or may not be feasible.

None exploits possible structure, such as the variance being a function (known, or unknown but smooth) of the mean.