Implementing GLS

Recall the assumptions of Approach 9:

\[ E(Y|\mathbf{x}) = f(\mathbf{x}, \boldsymbol{\beta}), \]
\[ \text{var}(Y|\mathbf{x}) = \sigma^2 g(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{x})^2. \]

Here:

- \( \boldsymbol{\beta} \) is, as before, the vector of parameters in the mean function;
- \( \boldsymbol{\theta} \) is a possible additional parameter in the structure of the variance function;
- \( \sigma^2 \) is an additional dispersion parameter.
**Ad hoc estimation scheme:**

1. Get initial estimate of $\beta$ using OLS;
2. Get initial estimate of $\theta$, if needed, and construct initial estimated weights
   \[
   \hat{w}_j = \frac{1}{g\left(\hat{\beta}_{\text{OLS}}, \hat{\theta}, x\right)^2}
   \]
3. Re-estimate $\beta$ using WLS, treating $\hat{w}_j$ as fixed: solve the estimating equation
   \[
   \sum_{j=1}^{n} \hat{w}_j \{ Y_j - f(x_j, \beta) \} f_{\beta}(x_j, \beta) = 0.
   \]
Digression

Gauss-Newton method for WLS (including OLS):

The equation

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

generally cannot be solved in closed form.

But if \( \beta^* \) is close to the solution \( \beta \),

\[
\begin{align*}
    f(x_j, \beta) &\approx f(x_j, \beta^*) + f_\beta(x_j, \beta^*)^T (\beta - \beta^*), \\
    f_\beta(x_j, \beta) &\approx f_\beta(x_j, \beta^*) + f_{\beta\beta}(x_j, \beta^*) (\beta - \beta^*)
\end{align*}
\]
Omitting terms likely to be small, the WLS estimating equation becomes

\[
\left\{ \sum_{j=1}^{n} w_j f_{\beta}(x_j, \beta^*) f_{\beta}(x_j, \beta^*)^T \right\} (\beta - \beta^*) 
\approx \sum_{j=1}^{n} w_j \left\{ Y_j - f(x_j, \beta^*) \right\} f_{\beta}(x_j, \beta^*)
\]
Write

\[
\mathbf{X}(\beta) = \begin{pmatrix}
    \mathbf{f}_\beta(x_1, \beta)^T \\
    \vdots \\
    \mathbf{f}_\beta(x_n, \beta)^T
\end{pmatrix}
\]

\[
\mathbf{f}(\beta) = \begin{pmatrix}
    \mathbf{f}(x_1, \beta) \\
    \vdots \\
    \mathbf{f}(x_n, \beta)
\end{pmatrix}
\]

\[
\mathbf{W} = \begin{pmatrix}
    w_1 & 0 & \ldots & 0 \\
    0 & w_2 & \ldots & 0 \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \ldots & w_n
\end{pmatrix}
\]
Then the approximate equation may be written

\[
\{ \mathbf{X}(\beta^*)^T \mathbf{WX}(\beta^*) \} (\beta - \beta^*) \approx \mathbf{X}(\beta^*)^T \mathbf{W} \{ \mathbf{Y} - f(\beta^*) \}
\]

or (if the inverse exists)

\[
\beta \approx \beta^* + \left\{ \mathbf{X}(\beta^*)^T \mathbf{WX}(\beta^*) \right\}^{-1} \mathbf{X}(\beta^*)^T \mathbf{W} \{ \mathbf{Y} - f(\beta^*) \}
\]

Iterative solution:

\[
\beta_{(a+1)} = \beta_{(a)} + \left\{ \mathbf{X}(\beta_{(a)})^T \mathbf{WX}(\beta_{(a)}) \right\}^{-1} \mathbf{X}(\beta_{(a)})^T \mathbf{W} \{ \mathbf{Y} - f(\beta_{(a)}) \}
\]

Note that \( \mathbf{W} \) is fixed throughout the inner circle of iteration.
Note that if the iteration converges, \( \beta_{(a)} \to \beta_{(\infty)} \), and 
\[
X(\beta_{(a)})^T WX(\beta_{(a)})
\] converges to a non-singular matrix, then 
\[
X(\beta_{(\infty)})^T W \{ Y - f(\beta_{(\infty)}) \} = 0
\]
which is the original estimating equation written in matrix form.

That is, the limit \( \beta_{(\infty)} \) does solve the original WLS problem.

This algorithm, suitably refined, is the default method in both R’s \texttt{nls()} and SAS’s \texttt{proc nlin}.
Instead of the nested iterations, we could solve the last equation directly.

**Note**

This is *not* equivalent to minimizing

\[
S_g(\beta) = \sum_{j=1}^{n} \frac{1}{g(\beta, \theta, x_j)^2} \left\{ Y_j - f(x_j, \beta) \right\}^2,
\]

because differentiating \( S_g(\beta) \) w.r.t. \( \beta \) also brings in the derivative of \( g(\cdot) \).
The equation can be solved by a Gauss-Newton method; a similar approximation leads to the iteration

\[
\beta_{(a+1)} = \beta_{(a)} + \left\{ X(\beta_{(a)})^T W_{(a)} X(\beta_{(a)}) \right\}^{-1} X(\beta_{(a)})^T W_{(a)} \{ Y - f(\beta_{(a)}) \}
\]

where the weight matrix \( W \) is now updated using the current value of \( \beta_{(a)} \).

That is, the weights are updated within the iteration, instead of being held constant, as in the WLS iteration; a nonlinear instance of Iteratively Reweighted Least Squares, IRWLS (note the redundancy!) or IWLS.
Estimating $\sigma^2$

By analogy with WLS, the natural estimator of $\sigma^2$ is either

$$\frac{1}{n} \sum_{j=1}^{n} \frac{\left\{ Y_j - f(x_j, \hat{\beta}_{GLS}) \right\}^2}{g(\hat{\beta}_{GLS}, \theta, x_j)^2}$$

or

$$\frac{1}{n - p} \sum_{j=1}^{n} \frac{\left\{ Y_j - f(x_j, \hat{\beta}_{GLS}) \right\}^2}{g(\hat{\beta}_{GLS}, \theta, x_j)^2}.$$
For fixed weights, the latter is unbiased.

For fixed weights and Gaussian $Y$, the former is ML and the latter is REML.

The latter is reported by most software.
Summary

Model

\[ E(Y|\mathbf{x}) = f(\mathbf{x}, \beta), \]
\[ \text{var}(Y|\mathbf{x}) = \sigma^2 g(\beta, \theta, \mathbf{x})^2. \]

Generalized least squares (GLS)

\[ \sum_{j=1}^{n} w_j \{ Y_j - f(\mathbf{x}_j, \beta) \} f_\beta(\mathbf{x}_j, \beta) = 0. \]
Motivation

- Loss function

\[ S_g(\beta) = \sum_{j=1}^{n} w_j \left\{ Y_j - f(x_j, \beta) \right\}^2, \]

with plugged-in weights \( w_j = g(\beta, \theta, x_j)^{-2} \).

- Gaussian distribution with mean \( f(x_j, \beta) \) and plugged-in weights \( w_j \).
Implementations

3-step GLS

1. Get initial estimate $\hat{\beta}^{(0)}$
2. Repeat until convergence, or for a fixed number of $C$ steps:
   1. Update the weights $\hat{w}_j = g(\hat{\beta}^{(t)}, x)^{-2}$
   2. Estimate $\beta$ using WLS:
      Repeat until convergence
      
      $$\beta_{(a+1)} = \beta_{(a)} + \left\{ X(\beta_{(a)})^T WX(\beta_{(a)}) \right\}^{-1} X(\beta_{(a)})^T W \{ Y - f(\beta_{(a)}) \}$$

3. Update the estimate $\hat{\beta}^{(t+1)}$
Comment

- There are 2 loops
- In the inner loop (step 2.2), the weight matrix $W$ is fixed
Implementations (continued)

Iteratively reweighted least squares (IRWLS)

1. Get initial estimate $\hat{\beta}^{(0)}$
2. Repeat until convergence:
   1. Update the weights $\hat{w}_j = g(\hat{\beta}^{(t)}, x)^{-2}$
   2. Estimate $\beta$ using WLS

\[ \beta_{(a+1)} = \beta_{(a)} + \left\{ X(\beta_{(a)})^T W_{(a)} X(\beta_{(a)}) \right\}^{-1} X(\beta_{(a)})^T W_{(a)} \left\{ Y - f(\beta_{(a)}) \right\} \]

3. Update the estimate $\hat{\beta}^{(t+1)}$
Comment

- There is only 1 loop
- The weight matrix $W$ is *iteratively updated*