Estimating $\theta$

Recall the general mean-variance specification

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

We have, so far, considered $\theta$ as a known constant.

In many cases, we instead need to estimate it.
One approach is to continue to use the Gaussian likelihood:

\[
-n \log \sigma - \sum_{j=1}^{n} \log g(\beta, \theta, x_j) - \frac{1}{2} \sum_{j=1}^{n} \frac{\{Y_j - f(x_j, \beta)\}^2}{\sigma^2 g(\beta, \theta, x_j)^2}.
\]

Differentiating w.r.t. $\theta$ leads to

\[
\sum_{j=1}^{n} \left[ \frac{\{Y_j - f(x_j, \beta)\}^2 - \sigma^2 g(\beta, \theta, x_j)^2}{\sigma^2 g(\beta, \theta, x_j)^2} \right] \nu_\theta(\beta, \theta, x_j) = 0.
\]
Here

$$\nu_{\theta}(\beta, \theta, x_j) = \frac{\partial \log g(\beta, \theta, x_j)}{\partial \theta} = \frac{g_{\theta}(\beta, \theta, x_j)}{g(\beta, \theta, x_j)}.$$ 

For consistency with earlier estimating equations, rewrite as

$$\sum_{j=1}^{n} \left[ \frac{\{Y_j - f(x_j, \beta)\}^2 - \sigma^2 g(\beta, \theta, x_j)^2}{2 \sigma^4 g(\beta, \theta, x_j)^4} \right] 2\sigma^2 g(\beta, \theta, x_j)^2 \nu_{\theta}(\beta, \theta, x_j) = 0.$$
Combining with the $\sigma^2$ equation derived earlier,

$$
\sum_{j=1}^{n} \left[ \frac{\{ Y_j - f(x_j, \beta) \}^2 - \sigma^2 g(\beta, \theta, x_j)^2}{2\sigma^4 g(\beta, \theta, x_j)^4} \right] 
\times \begin{pmatrix}
2\sigma g(\beta, \theta, x_j)^2 \\
2\sigma^2 g(\beta, \theta, x_j)^2 \nu(\beta, \theta, x_j)
\end{pmatrix} = 0.
$$
The $\sigma^2$ and $\theta$ equations may be put in the standard form

$$\sum_{j=1}^{n} D_j^T V_j^{-1} (s_j - m_j) = 0$$

We need

$$s_j = \{ Y_j - f (x_j, \beta) \}^2$$

$$m_j = \sigma^2 g (\beta, \theta, x_j)^2$$

$$D_j = \begin{pmatrix} 2\sigma g (\beta, \theta, x_j)^2 \\ 2\sigma^2 g (\beta, \theta, x_j)^2 \nu_\theta (\beta, \theta, x_j) \end{pmatrix}^T$$

$$V_j = 2\sigma^4 g (\beta, \theta, x_j)^4.$$
Since $\beta$ is also unknown, we must also solve the corresponding equation

$$\sum_{j=1}^{n} \frac{\{ Y_j - f(x_j, \beta) \} f_\beta(x_j, \beta)}{\sigma^2 g(\beta, \theta, x_j)^2}$$

$$+ \sum_{j=1}^{n} \left[ \frac{\{ Y_j - f(x_j, \beta) \}^2}{\sigma^2 g(\beta, \theta, x_j)^2} - 1 \right] \nu_\beta(\beta, \theta, x_j) = 0.$$ 

In principle, we must solve the $(q + 1)$ variance parameter equations and the $p \beta$ equations jointly.
Joint Estimating Equations

Gaussian ML:

$$\sum_{j=1}^{n} \begin{pmatrix} f_{\beta j} & 2\sigma^2 g_j^2 \nu_{\beta j} \\ 0 & 2\sigma^2 g_j^2 \left( \frac{1}{\sigma} \nu_{\theta j} \right) \end{pmatrix} \begin{pmatrix} \sigma^2 g_j^2 & 0 \\ 0 & 2\sigma^4 g^4_j \end{pmatrix}^{-1} \begin{pmatrix} Y_j - f_j \\ (Y_j - f_j)^2 - \sigma^2 g_j^2 \end{pmatrix} = 0.$$

GLS:

$$\sum_{j=1}^{n} \begin{pmatrix} f_{\beta j} & 0 \\ 0 & 2\sigma^2 g_j^2 \left( \frac{1}{\sigma} \nu_{\theta j} \right) \end{pmatrix} \begin{pmatrix} \sigma^2 g_j^2 & 0 \\ 0 & 2\sigma^4 g^4_j \end{pmatrix}^{-1} \begin{pmatrix} Y_j - f_j \\ (Y_j - f_j)^2 - \sigma^2 g_j^2 \end{pmatrix} = 0.$$
Remarks

Solve \((p + q + 1)\) estimating equations *jointly*.

Approach in GLS is called *pseudo-likelihood (PL)* approach; that is, some parameters are estimated by ML while others are estimated by ad hoc but consistent methods. — GLS-PL

When \(g(\cdot)\) does not depend on \(\beta\) but does involve unknown \(\theta\),

\[
\nu_\beta (\beta, \theta, x_j) = 0,
\]

and the normal theory ML and GLS equations coincide.
Implementation

Write estimating equations in this form, then solve using an idea similar to IRWLS

\[
\sum_{j=1}^{n} D_j^T(\alpha) V_j^{-1}(\alpha) \{ s_j(\alpha) - m_j(\alpha) \} = 0
\]

where

\[
D_j^T(\alpha) = \frac{\partial}{\partial \alpha} m_j(\alpha).
\]
Two strategies

Solve the entire set of parameters together:

\[ \alpha = \begin{pmatrix} \beta \\ \sigma \\ \theta \end{pmatrix}. \]

Iterate between:

- Given \( \alpha = \begin{pmatrix} \sigma \\ \theta \end{pmatrix} \), solve for \( \beta \);
- Given \( \beta \), solve for \( \alpha \).
Implementation

Use Taylor series approximation,

\[ \alpha \approx \alpha^* + \left\{ D^T(\alpha^*)V^{-1}(\alpha^*)D(\alpha^*) \right\}^{-1} D^T(\alpha^*)V^{-1}(\alpha^*) \{ s(\alpha^*) - m(\alpha^*) \} \]

where

\[ V(\alpha) = \text{block diag}\{ V_1(\alpha), \ldots, V_n(\alpha) \}, \]

and

\[ D^T(\alpha) = \{ D_1^T(\alpha), \ldots, D_n^T(\alpha) \} \]
Iterative update

\[ \alpha(a+1) = \alpha(a) + \left\{ D(a)^T V^{-1}(a) D(a) \right\}^{-1} D(a)^T V^{-1}(a) \{ s(a) - m(a) \} \]

Remark: The ML or PL quadratic equations are more unstable than are the linear equations, and can be ill-behaved in practice.
A General Class of Estimators of $\theta$

Motivation

Estimating equations for $\theta$ would be based on the residuals $\{Y_j - f(x_j, \beta)\}$.

For squared residuals, the effect of “outlying” or “unusual” observations can be magnified.

Other functions of the residuals might be considered.
Estimating Variance Parameters
Recall

- The standardized residual:
\[
\epsilon_j = \frac{Y_j - f(x_j, \beta)}{\sigma g(\beta, \theta, x_j)}
\]

- Box-Cox transformation:
\[
h(u, \lambda) = \begin{cases} 
  u^\lambda - 1 & \lambda \neq 0 \\
  \frac{\lambda}{\log |u|} & \lambda = 0
\end{cases}
\]

Key idea
Consider forming estimating equations based on general power transformations of absolute residuals.
Key assumption

The appropriate moments of $\epsilon_j$ are not dependent on $x_j$ and are constant for all $j$:

\[
E \left( |\epsilon_j|^\lambda \mid x_j \right) = E \left( |\epsilon_j|^\lambda \right) = \text{constant} \quad \forall j,
\]
\[
E \left( |\epsilon_j|^{2\lambda} \mid x_j \right) = E \left( |\epsilon_j|^{2\lambda} \right) = \text{constant} \quad \forall j.
\]

Definitions

Define $\eta$ such that

\[
e^{\lambda \eta} = \sigma^\lambda E(|\epsilon_j|^\lambda)
\]

Identify $|Y_j - f(x_j, \beta)|^{\lambda}$ as the “response”. 
It follows that

\[ E \left( \mid Y_j - f(x_j, \beta) \mid^\lambda \mid x_j \right) = \sigma^\lambda g(\beta, \theta, x_j)^\lambda E \left( \mid \epsilon_j \mid^\lambda \right) = e^{\lambda \eta} g(\beta, \theta, x_j)^\lambda \]

and

\[
\text{var} \left( \mid Y_j - f(x_j, \beta) \mid^\lambda \mid x_j \right) = \sigma^{2\lambda} g(\beta, \theta, x_j)^{2\lambda} \text{var} \left( \mid \epsilon_j \mid^\lambda \right) \\
= \sigma^{2\lambda} g(\beta, \theta, x_j)^{2\lambda} \left\{ E \left( \mid \epsilon_j \mid^{2\lambda} \right) - E \left( \mid \epsilon_j \mid^\lambda \right)^2 \right\} \\
= \left\{ \sigma^{2\lambda} E \left( \mid \epsilon_j \mid^{2\lambda} \right) - e^{2\lambda \eta} \right\} g(\beta, \theta, x_j)^{2\lambda}
\]
Estimating equations

\[
\sum_{j=1}^{n} \left\{ \frac{|Y_j - f(x_j, \beta)|^\lambda - e^{\lambda \eta} g(\beta, \theta, x_j)^\lambda}{g(\beta, \theta, x_j)^{2\lambda}} \right\} \lambda e^{\lambda \eta} g(\beta, \theta, x_j)^\lambda \tau_\theta(\beta, \theta, x_j) = 0
\]

defines a family of estimating equations for \( \theta \), indexed by \( \lambda \).

The quadratic PL equation is obtained when \( \lambda = 2 \).
Extended quasi-likelihood

Recall that quasi-likelihood (QL) was an attempt to give a “distributional” justification for GLS-type estimation of $\beta$ in models in which the variance depends on $\beta$ through the mean, but there are no additional, unknown, variance parameters $\theta$ (that is, $\theta$ is known).

When $\theta$ is unknown, the idea here is to define a “scaled exponential family-like” “loglikelihood” to be used as a basis for joint estimation of $\beta$, $\sigma$, and $\theta$. 
Recall

The log quasi-likelihood function

\[ \ell_{QL}(\beta, \sigma; y) = \frac{1}{\sigma^2} \int_y^{\mu} \frac{y - u}{g(u)^2} du. \]

where \( \mu = f(x, \beta) \).

Estimate \( \beta \) and \( \sigma \) by maximizing

\[ \sum_{j=1}^{n} \ell_{QL}(\beta, \sigma; y_j). \]
Define the *log extended quasi-likelihood* function

\[ \ell_{\text{EQL}}(\beta, \sigma, \theta; y) = \frac{1}{\sigma^2} \int_{\mu}^{y} \frac{y - u}{g(u)^2} du - \frac{1}{2} \log \left\{ 2\pi \sigma^2 g^2(y, \theta) \right\} \]

where \( \mu = f(x, \beta) \).

Estimate \( \beta, \sigma, \) and \( \theta \) by maximizing

\[ \sum_{j=1}^{n} \ell_{\text{EQL}}(\beta, \sigma, \theta; y_j). \]