Nonlinear Regression

As mentioned in the text, least squares methods are applicable to any model that is linear in the parameters. So, for example,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2^2 + \beta_5 X_1 X_2 + \varepsilon$$

can be fitted by least squares methods. Some other models such as

$$Y_{ij} = \alpha \frac{P_i^\beta P_j^\gamma}{D_{ij}^\delta} \varepsilon_{ij}$$

are nonlinear in the parameters but can be transformed into a model linear in the parameters, often by taking logs. Some models, however, cannot be transformed into linearity, such as

$$Y_i = \frac{\beta_1}{1 + \exp(\beta_2 + \beta_3 X_i)} + \varepsilon_i,$$

and are therefore called essentially nonlinear models. A general equation for a nonlinear model can be given by

$$Y_i = f(\boldsymbol{\beta}, \mathbf{x}_i^{'}) + \varepsilon_i,$$

or in matrix form as

$$\mathbf{y} = \mathbf{f}(\boldsymbol{\beta}, \mathbf{X}) + \varepsilon.$$

Under the assumption of independent normal errors with mean zero and constant variance, we can write the likelihood of the model as

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma_\varepsilon^2) &= \frac{1}{(2\pi\sigma_\varepsilon^2)^{n/2}} \exp\left\{ -\frac{\sum_{i=1}^n \left[Y_i - f(\boldsymbol{\beta}, \mathbf{x}_i^{'})\right]^2}{2\sigma_\varepsilon^2} \right\} \\ &= \frac{1}{(2\pi\sigma_\varepsilon^2)^{n/2}} \exp\left[-\frac{1}{2\sigma_\varepsilon^2} S(\boldsymbol{\beta}) \right] \end{aligned}$$

where

$$S(\boldsymbol{\beta}) \equiv \sum_{i=1}^n \left[Y_i - f(\boldsymbol{\beta}, \mathbf{x}_i^{'}) \right]^2.$$

To obtain the parameter estimates we take derivatives of $S(\boldsymbol{\beta})$ giving

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\sum_{i=1}^{n}\left[Y_i - f(\boldsymbol{\beta}, \mathbf{x}_i^{'})\right]\frac{\partial f(\boldsymbol{\beta}, \mathbf{x}_i^{'})}{\partial \boldsymbol{\beta}}.$$

We then set the derivatives to zero and solve to obtain the estimator $\mathbf{b}$. These estimating equations can be written in matrix form as

$$[\mathbf{F}(\mathbf{b}, \mathbf{X})]'\,[\mathbf{y} - \mathbf{f}(\mathbf{b}, \mathbf{X})] = \mathbf{0},$$

where $\mathbf{F}(\mathbf{b}, \mathbf{X})$ is the matrix of derivatives with i, j^{th} entry

$$F_{ij} = \frac{\partial f(\boldsymbol{b}, \mathbf{x}_i^{'})}{\partial B_j}.$$

The solution $\mathbf{b}$ is the MLE of $\boldsymbol{\beta}$, or if there are multiple solutions then we select the one with the smallest value of $S(\boldsymbol{\beta})$.

Solving the Estimating Equations

In general we cannot find an analytical solution to minimizing $S(\boldsymbol{\beta})$, so we use numerical methods. The text presents three methods: the steepest descent method, the Gauss-Newton method, and the Marquardt method. All of these methods begin with an initial starting value $\mathbf{b}^{(0)}$ and then use an updating algorithm to converge to the solution. They also all use the gradient function $\mathbf{d}(\mathbf{b}) = \partial S(\boldsymbol{b})/\partial \boldsymbol{b}$. With the steepest descent method, the updated value is

$$\mathbf{b}^{(l+1)} = \mathbf{b}^{(l)} - M_l\,\mathbf{d}(\mathbf{b}^{(l)})$$

where the value M_l is chosen so that $S(\boldsymbol{b}^{(l+1)}) < S(\boldsymbol{b}^{(l)})$. The above updating rule can be rewritten as

$$\mathbf{b}^{(l+1)} = \mathbf{b}^{(l)} + M_l\mathbf{F}_l'\mathbf{e}^{(l)}$$

where $\mathbf{F}_l = \mathbf{F}(\mathbf{b}^{(l)}, \mathbf{X})$ and $\mathbf{e}^{(l)} = \mathbf{y} - \mathbf{f}(\mathbf{b}^{(l)}, \mathbf{X})$ and the constant 2 above gets absorbed into the M_l value. The Gauss-Newton method uses a first-order Taylor expansion of $S(\boldsymbol{b})$ around $S(\boldsymbol{b}^{(l)})$ to get

$$\mathbf{b}^{(l+1)} = \mathbf{b}^{(l)} + M_l(\mathbf{F}_l'\mathbf{F}_l)^{-1}\mathbf{F}_l'\mathbf{e}^{(l)},$$

where again M_l is chosen so that $S(\boldsymbol{b}^{(l+1)}) < S(\boldsymbol{b}^{(l)})$. The Marquardt method is defined as

$$\mathbf{b}^{(l+1)} = \mathbf{b}^{(l)} + (\mathbf{F}_l'\mathbf{F}_l + M_l\mathbf{I}_p)^{-1}\mathbf{F}_l'\mathbf{e}^{(l)},$$

where M_0 is a small number and continues to decrease if $S(\boldsymbol{b}^{(l+1)}) < S(\boldsymbol{b}^{(l)})$ but increases if $S(\boldsymbol{b}^{(l+1)}) > S(\boldsymbol{b}^{(l)})$. The Marquardt method thus is adaptive and approaches the Gauss-Newton method when M_l gets small, but approaches the steepest descent method when M_l gets large.

The estimated sampling covariance matrix is obtained via the maximum likelihood approach and is

$$\widehat{V}(\mathbf{b}) = S_E^2 \left\{ [\mathbf{F}(\mathbf{b}, \mathbf{X})]' [\mathbf{F}(\mathbf{b}, \mathbf{X})] \right\}^{-1},$$

where the residuals $\mathbf{e} = \mathbf{y} - \mathbf{f}(\mathbf{b}, \mathbf{X})$ are used to estimate the error variance by

$$S_E^2 = \frac{\mathbf{e}'\mathbf{e}}{n-p}.$$