Gauss-Newton Regression
Econ 522a

Imagine an additive error regression model of the form

$$Y_t = h(X_t, \beta) + u_t,$$

where β is a K x 1 parameter vector to be estimated, $Cov\left[u_t, h(X_t, \beta)\right] = 0$, and u_t is i.i.d. with mean zero. In particular consider models in which $h(\cdot)$ is not a linear function of the parameters.

Nonlinear least squares (NLS) minimizes the sum of squared residuals:

$$S(\hat{\beta}^{nls}) = \left(\frac{1}{2}\right) \sum_{t=1}^{T} \left[Y_t - h(X_t, \hat{\beta}^{nls})\right]^2,$$

where $\left(\frac{1}{2}\right)$ is scaling factor that does not affect the solution. The corresponding F.O.C. are

$$\frac{\partial S(\hat{\beta}^{nls})}{\partial \hat{\beta}^{nls}} = -\sum_{t=1}^{T} \left[Y_t - h(X_t, \hat{\beta}^{nls})\right] \frac{\partial h(X_t, \beta)}{\partial \beta}|_{\hat{\beta}^{nls}} = 0$$

which produces a set of nonlinear normal equations that have to be solved by nonlinear optimization routines. The Gauss-Newton method is a very common optimization routine that is used for NLS. Essentially, this method iteratively minimizes the sums of squared residuals from linear approximations to the model.

A linear approximation to the nonlinear model follows from a first-order Taylor series approximation to $h(X_t, \beta)$ evaluated at some particular parameter vector value β^0:

$$h(X_t, \beta) \approx h(X_t, \beta^0) + \sum_{k=1}^{K} \frac{\partial h(X_t, \beta)}{\partial \beta_k}|_{\beta^0} \left(\beta_k - \beta_k^0\right).$$

Therefore,

$$h(X_t, \beta) = h(X_t, \beta^0) + \sum_{k=1}^{K} \frac{\partial h(X_t, \beta)}{\partial \beta_k}|_{\beta^0} \left(\beta_k - \beta_k^0\right) + R_t^0,$$

where R_t^0 is the remainder term, i.e.

$$R_t^0 = h(X_t, \beta) - h(X_t, \beta^0) - \sum_{k=1}^{K} \frac{\partial h(X_t, \beta)}{\partial \beta_k}|_{\beta^0} \left(\beta_k - \beta_k^0\right).$$

Notice that we can express this linearized model as a conventional regression model:

$$Y_t = h(X_t, \beta^0) + \sum_{k=1}^{K} \frac{\partial h(X_t, \beta)}{\partial \beta_k}|_{\beta^0} \left(\beta_k - \beta_k^0\right) + R_t^0 + u_t$$

$\Rightarrow$

$$Y_t - h(X_t, \beta^0) + \sum_{k=1}^{K} \frac{\partial h(X_t, \beta)}{\partial \beta_k}|_{\beta^0} \ \beta_k^0 = \sum_{k=1}^{K} \frac{\partial h(X_t, \beta)}{\partial \beta_k}|_{\beta^0} \ \beta_k + R_t^0 + u_t$$

or

$$Y_t^0 = \sum_{k=1}^{K} \frac{\partial h(X_t, \beta)}{\partial \beta_k}|_{\beta^0} \ \beta_k + u_t^0,$$

where $Y_t^0 = Y_t - h(X_t, \beta^0) + \sum_{k=1}^{K} \frac{\partial h(X_t, \beta)}{\partial \beta_k}|_{\beta^0} \ \beta_k^0$, $u_t^0 = R_t^0 + u_t$, and $\frac{\partial h(X_t, \beta)}{\partial \beta_k}|_{\beta^0}$ is the kth regressor. We can express the linearized regression model in matrix form:

$$Y^0 = H^0 \beta + u^0, \tag{1}$$

where Y^0 and u^0 are T x 1 observation vectors corresponding to Y_t^0 and u_t^0, and H^0 is the T x K observation matrix corresponding to the K regressors given by $\frac{\partial h(X_t, \beta)}{\partial \beta_k}|_{\beta^0}$, $k = 1, ..., K$. Our first round estimate of β is obtained from

$$\hat{\beta}^{(1)} = \left(H^{0\prime} H^0 \right)^{-1} H^{0\prime} Y^0.$$

Continuing onto the next iteration, we run the regression

$$Y^{(1)} = H^{(1)} \beta + u^{(1)},$$

where the elements of $Y^{(1)}$ are obtained from

$Y_t^{(1)} = Y_t - h(X_t, \hat{\beta}^{(1)}) + \sum_{k=1}^{K} \frac{\partial h(X_t, \beta)}{\partial \beta_k}|_{\beta^{(1)}} \ \beta_k^{(1)}$ and $H^{(1)}$ is the T x K observation matrix corresponding to the K regressors given by $\frac{\partial h(X_t, \beta)}{\partial \beta_k}|_{\beta^{(1)}}$, $k = 1, ..., K$. The second round estimate of β is obtained from

$$\hat{\beta}^{(2)} = \left(H^{(1)\prime} H^{(1)} \right)^{-1} H^{(1)\prime} Y^{(1)}.$$

Subsequent iterations occur until $\hat{\beta}^{(j)} - \hat{\beta}^{(j-1)} \approx 0$ or $\frac{\hat{\beta}^{(j)} - \hat{\beta}^{(j-1)}}{\hat{\beta}^{(j-1)}} \approx 0$. If convergence is reached on the jth iteration, then $\hat{\beta}^{nls} = \hat{\beta}^{(j)}$. The estimated variance/covariance matrix for the NLS estimator is given by

$$\widehat{Var}(\hat{\beta}^{nls}) = \hat{\sigma}_u^2 \left(H^{nls\prime} H^{nls} \right)^{-1}$$

where $\hat{\sigma}_u^2 = \frac{\hat{u}^{(nls)\prime} \hat{u}^{(nls)}}{T \text{ or } T-K}$, $\hat{u}_t^{(nls)} = Y_t - h(X_t, \hat{\beta}^{nls})$, and $H^{nls} = H^{(j)} \approx H^{(j-1)}$.

An equivalent way to express the Gauss-Newton non-linear method is to subtract $H^0\beta^0$ from both sides of (1) to obtain

$$\Delta Y^0 = H^0 \Delta\beta^{(1)} + u^0 \tag{2}$$

where $\Delta Y^0 = Y^0 - H^0\beta^0$, i.e. $\Delta Y_t^0 = Y_t - h(X_t, \beta^0)$, and $\Delta\beta^{(1)} = (\beta - \beta^0)$. Now we could estimate (2) by OLS to obtain

$$\widehat{\Delta\beta}^{(1)} = \left(H^{0\prime}H^0\right)^{-1} H^{0\prime}\Delta Y^0.$$

Next, we back out our 1st round estimate of β from

$$\hat{\beta}^{(1)} = \widehat{\Delta\beta}^{(1)} + \beta^0.$$

Continuing onto the next iteration, we run the regression

$$\Delta Y^{(1)} = H^{(1)}\Delta\beta^{(2)} + u^{(1)},$$

where $\Delta Y_t^{(1)} = Y_t - h(X_t, \hat{\beta}^{(1)})$ and $H^{(1)}$ is the T x K observation matrix corresponding to the K regressors given by $\frac{\partial h(X_t,\beta)}{\partial \beta_k}|_{\beta^{(1)}}$, $k = 1, ..., K$. The second round estimate of β is obtained from

$$\hat{\beta}^{(2)} = \widehat{\Delta\beta}^{(2)} + \hat{\beta}^{(1)}.$$

One can see that this is an iterative process, so on the jth iteration we have an updated estimate

$$\hat{\beta}^{(j)} = \widehat{\Delta\beta}^{(j)} + \hat{\beta}^{(j-1)}.$$

Convergence is reached when $\widehat{\Delta\beta}^{(j)} \approx 0$ or $\frac{\widehat{\Delta\beta}^{(j)}}{\hat{\beta}^{j-1}} \approx 0$ according to some numerical criterion, e.g. $\widehat{\Delta\beta}^{(j)} < 10^{-5}$.

The estimated asymptotic variance covariance matrix for $\hat{\beta}^{nls}$ is estimated by

$$\widehat{Var}\left(\hat{\beta}^{nls}\right) = \hat{\sigma}_u^2 \left(H^{nls\prime}H^{nls}\right)^{-1},$$

where $\hat{\sigma}_u^2 = \frac{\sum_{t=1}^T \left[Y_t - h(X_t, \hat{\beta}^{nls})\right]^2}{T - K \text{ or } T}$ and H^{nls} is the T x K observation matrix corresponding to the K regressors given by $\frac{\partial h(X_t,\beta)}{\partial \beta_k}|_{\hat{\beta}^{nls}}$, $k = 1, ..., K$.