5. The Newton-Raphson Algorithm

5.1 The nonlinear least squares estimates - The single-parameter case

$y = f(X, \beta) + e$, $\mathrm{E}[e] = 0$ e $E[ee'] = \sigma^2 I$.

We want to find the value of β that minimizes

$$S(\beta) = [y - f(X, \beta)]'[y - f(X, \beta)] = \sum_{t=1}^{T} [y_t - f(x_t, \beta)]^2.$$

Examples

Compute the least squares estimator for the parameter β in the following models:

1. The linear regression model,

$$y_t = \beta x_t + e_t, \quad t = 1, ..., T.$$

2. The non-linear regression model, $y_t = f(x_t, \beta) + e_t$ where

   $f(x, \beta) = \beta x_{t1} + \beta^2 x_{t2}$, that is,

$$y_t = \beta x_{t1} + \beta^2 x_{t2} + e_t, \quad t = 1, ..., T.$$

5. The Newton-Raphson Algorithm

$$y = f(X, \beta) + e, \ \mathrm{E}[e] = 0 \text{ e } E[ee'] = \sigma^2 I.$$

We want to find the value of β that minimizes

$$S(\beta) = [y - f(X, \beta)]'[y - f(X, \beta)] = \sum_{t=1}^{T} [y_t - f(x_t, \beta)]^2.$$

We replace $S(\beta)$ with a second-order Taylor series approximation, namely,

$$S(\beta) \simeq S(\beta_1) + \left.\frac{dS}{d\beta}\right|_{\beta_1} (\beta - \beta_1) + \frac{1}{2} \left.\frac{d^2 S}{d\beta^2}\right|_{\beta_1} (\beta - \beta_1)^2.$$

Our aim is to find what value of β minimizes $S(\beta)$ given that we begin with some know initial value β_1. Differentiating with respect to β, and using the notation,

$$h(\beta_1) = \frac{d^2 S}{d\beta^2}\Big|_{\beta_1}$$

yields,

$$\frac{dS}{d\beta} \simeq \frac{dS}{d\beta}\Big|_{\beta_1} + h(\beta_1) \cdot (\beta - \beta_1)$$

Setting this derivative equal to 0 and solving for β leads to a second value for β, say β_2, which is given by

$$\beta_2 = \beta_1 - (h(\beta_1))^{-1} \frac{dS}{d\beta}\Big|_{\beta_1}.$$

Continuing this procedure leads to the $(n+1)$th value for β being given by

$$\beta_{n+1} = \beta_n - (h(\beta_n))^{-1} \left.\frac{dS}{d\beta}\right|_{\beta_n}.$$

- If the process converges in the sense that $\beta_{n+1} - \beta_n$ converges to zero, then it must be true that $\left.\frac{dS}{d\beta}\right|_{\beta_n} = 0$, the necessary condition for a minimum (or a maximum).
- If $h(\beta_1)$ is positive we will go in the right direction (toward a minimum) if β_1 is sufficiently close to the minimizing value.
- Recall that if $h(\beta_1)$ is positive then h still be positive in a neighborhood of a minimum.
- To establish whether a particular minimum is local or global, a number of different starting values need to be tried.

5.2 The nonlinear least squares estimates - The general K parameter case

For the nonlinear model
$y = f(X, \beta) + e$ where β is a K-dimensional vector of *unknown* parameters,E[e] $= 0$ e $E[ee'] = \sigma^2 I$. the nth iteration of the Newton-Raphson algorithm designed to find that value of β that minimizes $S(\beta) = ee'$ is given by

$$\beta_{n+1} = \beta_n - H_n^{-1} \left.\frac{\partial S}{\partial \beta}\right|_{\beta_n}$$

where

$$\left.\frac{\partial S}{\partial \beta}\right|_{\beta_n} = \left(\frac{\partial S}{\partial \beta_1}, \frac{\partial S}{\partial \beta_2}, \ldots, \frac{\partial S}{\partial \beta_K}\right)' \Big|_{\beta_n}$$

is the gradient vector evaluated at β_n.

H_n is the $(K \times K)$ Hessian matrix evaluated at β_n,

$$H_n = \frac{\partial^2 S}{\partial \beta \, \partial \beta'}\bigg|_{\beta_n} = \begin{bmatrix} \frac{\partial^2 S}{\partial \beta_1^2} & \cdots & \frac{\partial^2 S}{\partial \beta_1 \, \partial \beta_K} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 S}{\partial \beta_K \, \partial \beta_1} & \cdots & \frac{\partial^2 S}{\partial \beta_k^2} \end{bmatrix}_{(\beta_n)}$$

- The algorithm will lead to the right direction (toward a minimum) from the point β_1 if H_n is definite positive.

To compute the vector of parameters that maximizes the likelihood function, $L(\beta, x)$ we can apply the Raphson method,

for the maximum likelihood estimates:

$$\beta_{n+1} = \beta_n - \left[\frac{\partial^2 L}{\partial\beta\,\partial\beta'}\right]^{-1}\Big|_{\beta_n} \frac{\partial L}{d\beta}\Big|_{\beta_n}.$$