Regressão não linear

Considere um modelo geral:

$$Y_t = f\left(X_t, \beta\right) + \varepsilon_t \qquad T \text{ observações}$$

onde

X_t é um vetor de variáveis exógenas (Nx1)

β é um vetor de parâmetros (Kx1)

Y_t é a variável dependente

ε_t é o erro aleatório

K e N não são necessariamente iguais nesse caso.

Exemplo:

Considere a função de produção Cobb-Douglas $Q_t = \alpha L_t^{\beta_1} K_t^{\beta_2} + \varepsilon_t$

$$Q_t = Y_t \qquad X_t = \begin{bmatrix} L_t \\ K_t \end{bmatrix} \qquad \beta = \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

$$f\left(X_t, \beta\right) = \alpha L_t^{\beta_1} K_t^{\beta_2}$$

Considere um modelo simples com um só parâmetro.

$$Y_t = f\left(X_t, \beta\right) + \varepsilon_t = \beta X_{t1} + \beta^2 X_{t2} + \varepsilon_t$$

A estimativa de mínimos quadrados não linear de β é definida como o valor de β que minimiza a soma de quadrados dos erros.

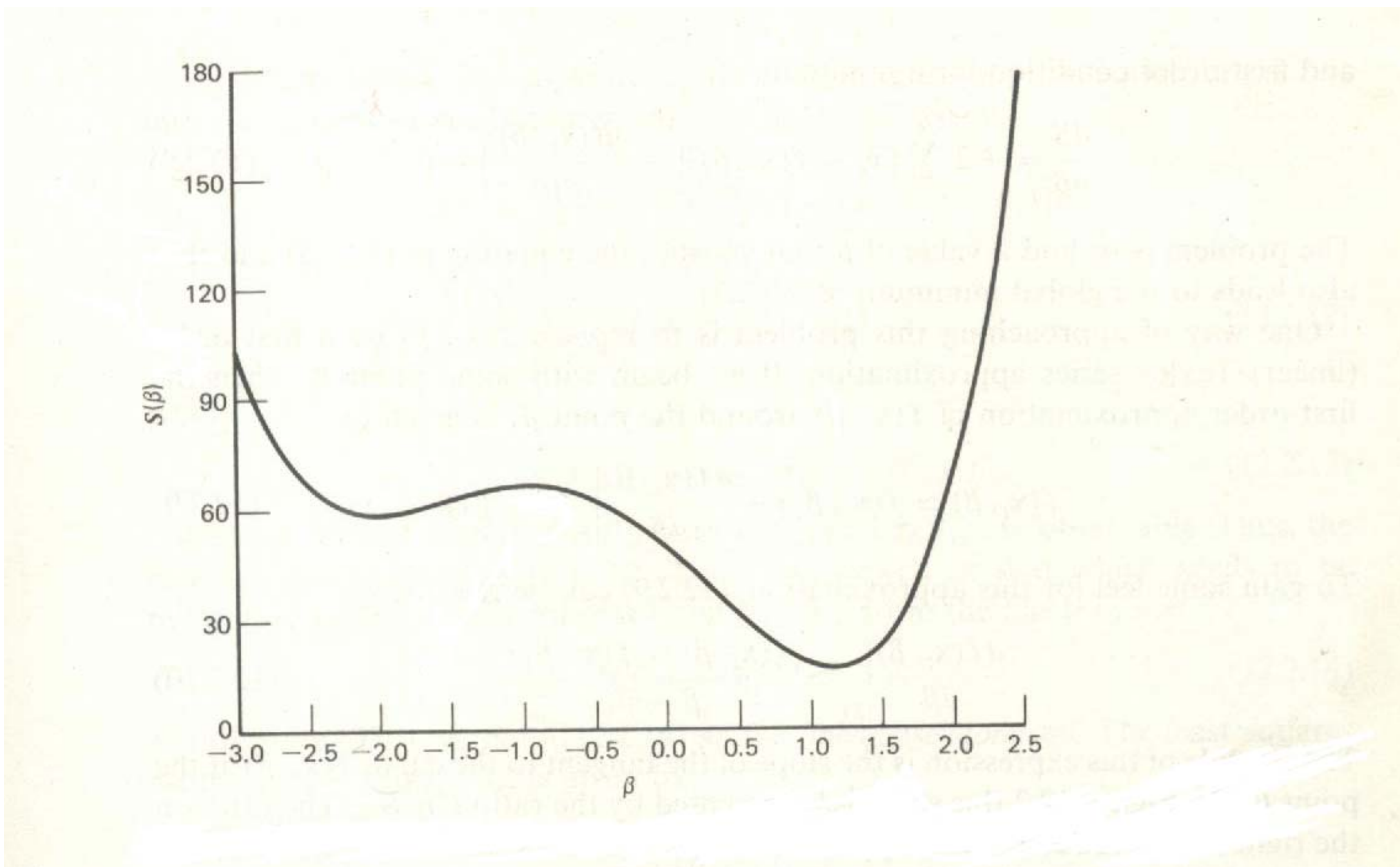
$$S(\beta) = \sum_{t=1}^{T} \varepsilon_t^2 = \sum_{t=1}^{T} \left[Y_t - f\left(X_t, \beta\right)\right]^2 = \sum_{t=1}^{T} \left(Y_t - \beta X_{t1} - \beta^2 X_{t2}\right)^2$$

Condição de 1ª. ordem

$$\frac{\partial S}{\partial \beta} = 2\sum_{t=1}^{T} \left(Y_t - f\left(X_t, \beta\right)\right)\left(-\frac{\partial f\left(X_t, \beta\right)}{\partial \beta}\right) = 2\sum \left(Y_t - \beta X_{t1} - \beta^2 X_{t2}\right)\cdot\left(-X_{t1} - 2\beta X_{t2}\right) = 0$$

Rearranjando,

$$2\beta^3 \sum X_{t2}^2 + 3\beta^2 \sum X_{t1} X_{t2} + \beta\left(\sum X_{t1}^2 - 2\sum X_{t2} Y_t\right) - \sum X_{t1} Y_t = 0$$



Para achar o valor de β que satisfaça a equação

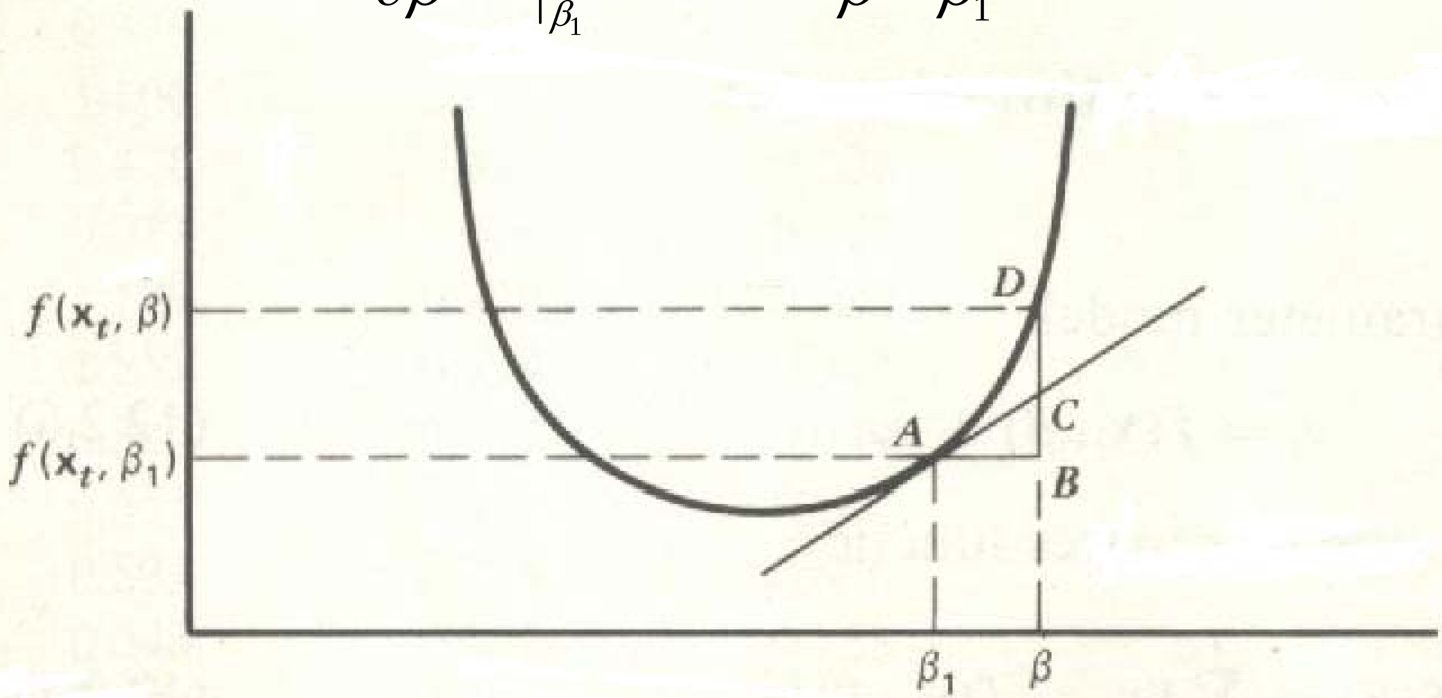
$$\frac{\partial S}{\partial \beta} = 2\sum\left(y_t - f\left(X_t, \beta\right)\right)\left(-\frac{\partial f\left(X_t, \beta\right)}{\partial \beta}\right) = 0$$

e forneça um mínimo global, vamos substituir $f(X_t, \beta)$ pela aproximação da série de Taylor de 1ª. ordem. Começando com um ponto inicial β_1 ,a aproximação de 1ª. ordem de $f(X_t, \beta)$ em torno do ponto β_1 é:

$$f\left(X_t, \beta\right) \cong f\left(X_t, \beta_1\right) + \left.\frac{\partial f\left(X_t, \beta\right)}{\partial \beta}\right|_{\beta_1} \left(\beta - \beta_1\right)$$

O que significa esta aproximação?

$$\left.\frac{\partial f\left(X_t,\beta\right)}{\partial \beta}\right|_{\beta_1} \cong \frac{f\left(X_t,\beta\right)-f\left(X_t,\beta_1\right)}{\beta-\beta_1}$$



O lado esquerdo da expressão é a inclinação da reta tangente à curva $f(X_t, \beta)$ no ponto β_1=CB/BA. O lado direito, DB/BA, pode ser visto como a inclinação da reta unindo os pontos D e A, e esta inclinação sendo usada para aproximar a inclinação CA.

Considere,

$$Z_t(\beta) = \frac{\partial f(X_t, \beta)}{\partial \beta} \quad \text{e} \quad Z_t(\beta_1) = \left.\frac{\partial f(X_t, \beta)}{\partial \beta}\right|_{\beta_1}$$

$$S(\beta) = \sum \left(Y_t - f(X_t, \beta)\right)^2$$

Substituindo

$$f(X_t, \beta) \cong f(X_t, \beta_1) + Z_t(\beta_1)(\beta - \beta_1)$$

em $S(\beta)$ fica:

$$S(\beta) = \sum \left(Y_t - f(X_t, \beta_1) - Z_t(\beta_1)(\beta - \beta_1)\right)^2$$

$$= \sum \left(Y_t - f(X_t, \beta_1) + Z_t(\beta_1) \cdot \beta_1 - Z_t(\beta_1) \cdot \beta\right)^2$$

$$= \sum \left(\overline{Y}_t(\beta_1) - Z_t(\beta_1) \cdot \beta\right)^2$$

onde

$$\overline{Y}_t(\beta_1) = Y_t - f(X_t, \beta_1) + Z_t(\beta_1) \cdot \beta_1$$

Observe que para um dado valor de β_1, $\overline{Y}_t(\beta_1)$ e $Z_t(\beta_1)$ são conhecidos. Portanto, a soma de quadrado dos resíduos $S(\beta)$ acima, pode ser vista como aquela que precisa ser minimizada para achar a estimativa de mínimos quadrados para β do modelo linear.

$$\overline{Y}_t(\beta_1) = Z_t(\beta_1) \cdot \beta + \varepsilon_t$$

que é conhecido como pseudo-modelo.

A estimativa de mínimos quadrados é:

$$\beta_2 = \frac{\sum \overline{Y}_t(\beta_1) Z_t(\beta_1)}{\sum Z_t(\beta_1)^2} = \left[Z(\beta_1)' Z(\beta_1) \right]^{-1} Z(\beta_1)' \overline{Y}(\beta_1)$$

Este processo pode ser repetido.

$$\bar{Y}(\beta_2) = Z(\beta_2) \cdot \beta + \varepsilon$$

$$\beta_3 = \left[Z(\beta_2)' Z(\beta_2) \right]^{-1} Z(\beta_2)' \bar{Y}(\beta_2)$$

$$\ldots$$

$$\beta_{n+1} = \left[Z(\beta_n)' Z(\beta_n) \right]^{-1} Z(\beta_n)' \bar{Y}(\beta_n) =$$

$$\left[Z(\beta_n)' Z(\beta_n) \right]^{-1} Z(\beta_n)' \left(Y - f(X, \beta_n) + Z(\beta_n)\beta_n\right)$$

$$\beta_{n+1} = \beta_n + \left[Z(\beta_n)' Z(\beta_n) \right]^{-1} Z(\beta_n)' \left[Y - f(X, \beta_n)\right]$$

O processo caminha em direção a um mínimo.

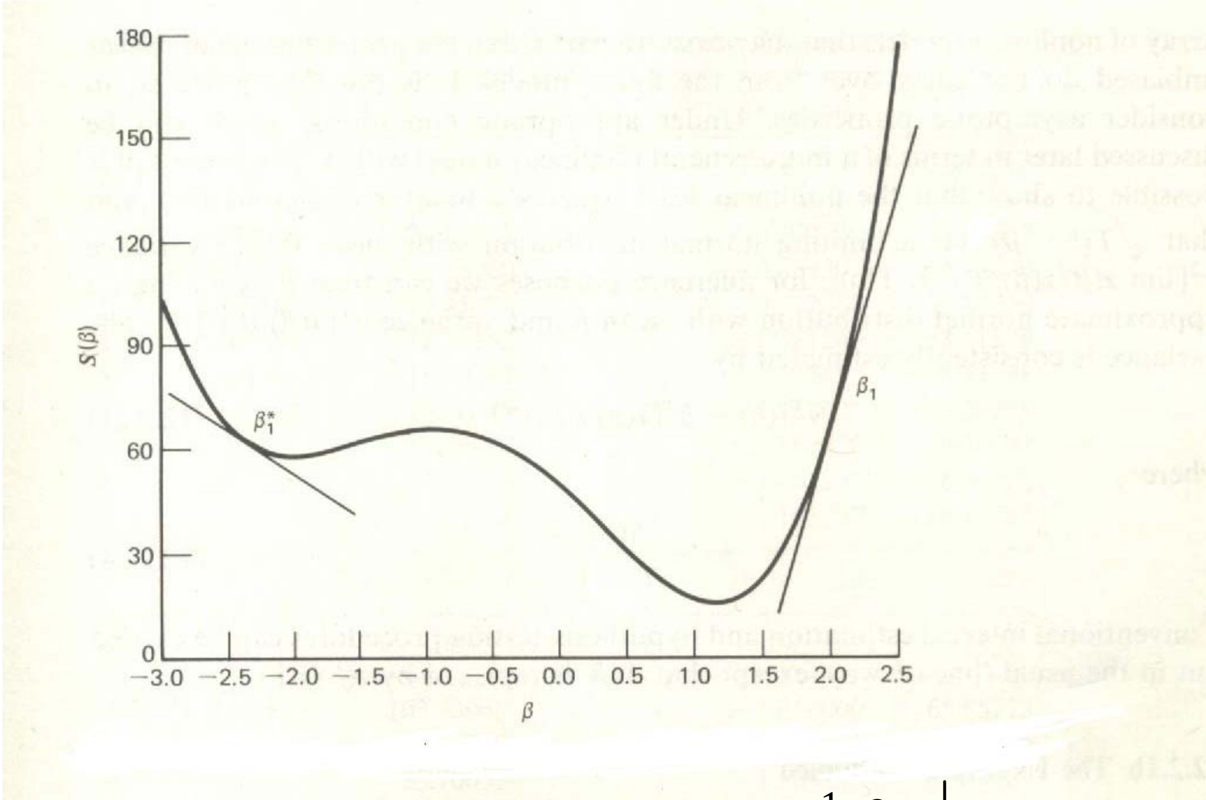
Lembre-se que

$$\frac{\partial S}{\partial \beta} = -2Z(\beta)'\left[Y - f(X_t, \beta)\right]$$

então

$$\beta_{n+1} = \beta_n - \frac{1}{2}\left[Z(\beta_n)' Z(\beta_n)\right]^{-1} \left.\frac{\partial S}{\partial \beta}\right|_{\beta_n}$$

Se $\beta_{n+1} = \beta_n$, então $\partial S/\partial\beta = 0$ e, portanto, β_n satisfaz a condição necessária para mínimo.



$$\beta_{n+1} = \beta_n - \frac{1}{2}\left[Z(\beta_n)' Z(\beta_n) \right]^{-1} \left. \frac{\partial S}{\partial \beta} \right|_{\beta_n}$$

No ponto β_1 , $\partial S/\partial \beta$ é positivo. Como $Z(\beta)'Z(\beta)$ é uma soma de quadrados, $[Z(\beta)'Z(\beta)]^{-1}$ será sempre positivo. Portanto, o 2º. valor de β será menor do que o 1º.

No ponto β^*_1 , $\partial S/\partial \beta$ é negativo e mudança em β será em direção ao mínimo.

Convergência em probabilidade.

Seja $\hat{\theta}_T$ o estimador de θ baseado em uma amostra de tamanho *T*.

Então $\hat{\theta}_T$ converge em probabilidade para θ se:

$$\lim_{T\to\infty}\left(\operatorname{Prob}\left|\hat{\theta}_T - \theta\right| < \varepsilon\right) = 1 \qquad \Rightarrow \qquad \lim_{T\to\infty}\left(\operatorname{Prob}(\theta - \varepsilon < \hat{\theta}_T < \theta + \varepsilon)\right) = 1$$

Se $\hat{\theta}_T$ converge em probabilidade para θ ⇒ plim $\hat{\theta}_T = \theta$

Um estimador $\hat{\theta}_T$ do parâmetro θ é consistente se e somente se plim $\hat{\theta}_T = \theta$

Propriedades assintóticas do estimador:

b é um estimador consistente com distribuição aproximadamente normal com média β e variância

$$\sigma^2 \left[Z(\beta)' Z(\beta) \right]^{-1}$$

$$Va\hat{r}(b) = s^2 \left[Z(b)' Z(b) \right]^{-1}$$

$$s^2 = \frac{s(b)}{T - K}$$

Voltando ao exemplo: $$Y_t = \beta X_{t1} + \beta^2 X_{t2} + \varepsilon_t$$

$$Z(\beta) = \frac{\partial f(X,\beta)}{\partial \beta} = X_1 + 2\beta X_2$$

$$Z(\beta)' Z(\beta) = \sum_{t=1}^{T} \left(X_{1t} + 2\beta X_{2t}\right)^2$$

$$Z(\beta)' \left[Y - f(X,\beta)\right] = \sum \left(X_{1t} + 2\beta X_{2t}\right)\left(Y_t - \beta X_{1t} - \beta^2 X_{2t}\right)$$

$$\beta_{n+1} = \beta_n + \frac{\sum \left(X_1 + 2\beta_n X_2\right)\left(Y - \beta_n X_1 - \beta_n^2 X_2\right)}{\sum \left(X_1 + 2\beta_n X_2\right)^2}$$

Se β_0=1, então

$$\beta_1 = 1 + \frac{\sum X_1 Y - \beta \sum X_1^2 - \beta^2 \sum X_1 X_2 + 2\beta \sum X_2 Y - 2\beta^2 \sum X_1 X_2 - 2\beta^3 \sum X_2^2}{\sum X_1^2 + 4\beta \sum X_1 X_2 + 4\beta^2 \sum X_2^2}$$

Considere os dados

X_1	X_2	Y	X_1Y	X_1^2	X_2Y	X_2^2	X_1X_2
0	3	11	0	0	33	9	0
1	0	3	3	1	0	0	0
2	3	17	34	4	51	9	6
3	0	5	15	9	0	0	0
4	1	12	48	16	12	1	4
total			100	30	96	19	10

$$\hat{\beta}_1 = 1 + \frac{100 - 30 - 10 + 2 \cdot 96 - 2 \cdot 10 - 2 \cdot 19}{30 + 4 \cdot 10 + 4 \cdot 19} = 1 + \frac{194}{146} = 1 + 1,33$$

$$\hat{\beta}_1 = 2,33$$

$$\hat{\beta}_2 = 2,33 + \ldots$$

converge para $\hat{\beta} = 2$

$$\hat{Y} = \hat{\beta} X_1 + \hat{\beta}^2 X_2 = 2X_1 + 4X_2$$

$$e = Y - \underbrace{\left(2X_1 + 4X_2\right)}_{\hat{Y}}$$

Y	$\hat{Y}$	e
11	12	–1
3	2	1
17	16	1
5	6	–1
12	12	0

$$s^2 = \frac{\sum e^2}{4} = \frac{4}{4} = 1$$

$$Var(b) = \left(Z(\beta)' Z(\beta) \right)^{-1} s^2 = \left[\sum \left(X_1 + 4X_2 \right)^2 \right]^{-1} .1 = [414]^{-1} \cdot 1 = 0{,}00242$$

<u>Forma Matricial</u>

O modelo

$$Y_t = f\left(X_t, \beta\right) + \varepsilon_t$$

onde β é um vetor($K\times 1$), na forma matricial fica,

$$Y = f\left(X, \beta\right) + \varepsilon$$

Pressupondo que E(ε)=0 e E($\varepsilon\varepsilon$')=$\sigma^2 I$, a estimativa de mínimos quadrados não linear do vetor β é o valor de β que minimiza a soma de quadrado dos erros.

$$S\left(\beta\right) = \varepsilon'\varepsilon = \left[Y - f\left(X, \beta\right)\right]'\left[Y - f\left(X, \beta\right)\right] = \sum_{t=1}^{T}\left[Y_t - f\left(X_t, \beta\right)\right]^2$$

Neste caso, existem K condições de primeira ordem para o mínimo, definidas a partir de K derivadas,

$$\frac{\partial S}{\partial \beta} = -2\frac{\partial f\left(X, \beta\right)'}{\partial \beta}\left[Y - f\left(X, \beta\right)\right] = 0$$

onde $\partial f(X,\beta)'/\partial\beta$ é uma matriz (K×T) de derivadas onde o (K, t) ésimo elemento é $\partial f(X_t,\beta)'/\partial\beta_K$.

Sendo Z(β) o transposto da matriz $\dfrac{\partial f(X,\beta)'}{\partial\beta}$, tem-se

$$Z(\beta)=\frac{\partial f(X,\beta)}{\partial\beta'}=\begin{bmatrix}\dfrac{\partial f(X_1,\beta)}{\partial\beta_1} & \cdots & \dfrac{\partial f(X_1,\beta)}{\partial\beta_K}\\ \vdots & & \\ \dfrac{\partial f(X_T,\beta)}{\partial\beta_1} & \cdots & \dfrac{\partial f(X_T,\beta)}{\partial\beta_K}\end{bmatrix}_{T\times K}$$

A condição de 1ª. ordem para mínimo será.

$$Z(\beta)'\left[Y - f(X,\beta)\right] = 0$$

Aproximação de Taylor para a t-ésima observação é,

$$f(X_t,\beta) \cong f(X_t,\beta_1) + \left[\left.\frac{\partial f(X_t,\beta)}{\partial \beta_1}\right|_{\beta_1} \cdots \left.\frac{\partial f(X_t,\beta)}{\partial \beta_K}\right|_{\beta_1} \right](\beta - \beta_1)$$

Incluindo-se todas as *T* observações tem-se,

$$f(X,\beta) \cong f(X,\beta_1) + Z(\beta_1)(\beta - \beta_1)$$

então,

$$Y \cong f\left(X_t, \beta_1\right) + Z\left(\beta_1\right)\left(\beta - \beta_1\right) + \varepsilon$$

$$Y \cong f\left(X, \beta_1\right) + Z\left(\beta_1\right)\beta - Z\left(\beta_1\right)\beta_1 + \varepsilon$$

$$Y - f\left(X, \beta_1\right) + Z\left(\beta_1\right)\beta_1 \cong Z\left(\beta_1\right)\beta + \varepsilon$$

ou

$$\tilde{Y}\left(\beta_1\right) \cong Z\left(\beta_1\right)\beta + \varepsilon$$

é um pseudo-modelo, então

$$\beta_2 = \left[Z\left(\beta_1\right)' Z\left(\beta_1\right)\right]^{-1} \left[Z\left(\beta_1\right)' \tilde{Y}\left(\beta_1\right)\right]$$

$$\beta_2 = \left[Z\left(\beta_1\right)' Z\left(\beta_1\right)\right]^{-1} Z\left(\beta_1\right)' \left[Y - f\left(X, \beta_1\right) + Z\left(\beta_1\right)\beta_1\right]$$

$$\beta_2 = \beta_1 + \left[Z\left(\beta_1\right)' Z\left(\beta_1\right)\right]^{-1} Z\left(\beta_1\right)' \left[Y - f\left(X, \beta_1\right)\right]$$

Continuando o processo, a *n*-ésima iteração do algorítmo de Gauss-Newton é dada por:

$$\beta_{n+1} = \beta_n + \underbrace{\overbrace{\left[Z(\beta_n)' Z(\beta_n) \right]^{-1}}^{B} Z(\beta_n)' \left[Y - f(X, \beta_n) \right]}_{A}$$

Quando a iteração pára é porque $\hat{\beta}_{n+1} = \hat{\beta}_n$, isto é, A=0. Como B≠0, então $Z(\beta_n)' \left[Y - f(X_1 \beta_n) \right] = 0$ e esta é exatamente é exatamente a condição de 1ª. ordem para o mínimo.

$$\sum{}_b = \hat{\sigma}^2 \left[Z(b)' Z(b) \right]^{-1}$$

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}\hat{\varepsilon}'}{T - K}$$

Algorítimo de Newton-Raphson

$$Y = f(X, \beta) + \varepsilon$$

$$E(\varepsilon) = 0 \qquad E(\varepsilon\varepsilon') = \sigma^2 I$$

queremos encontrar o valor de β que minimiza,

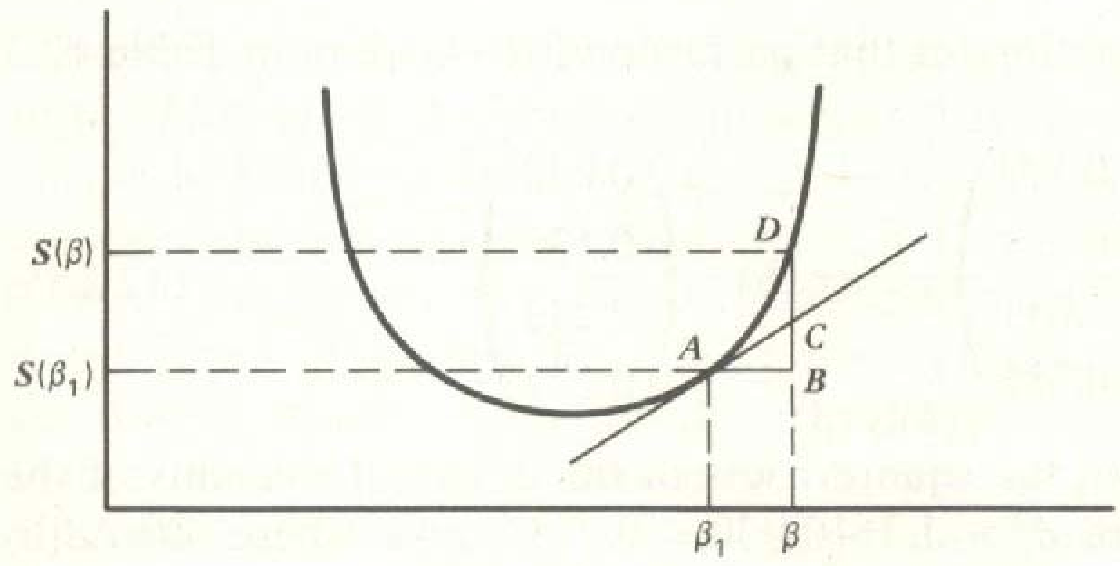
$$S(\beta) = [Y - f(X,\beta)]'[Y - f(X,\beta)] = \sum_{t=1}^{T} [Y_t - f(X_t,\beta)]^2$$

Iniciamos substituindo $S(\beta)$ pela aproximação de Taylor de 2ª. ordem.

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

O que significa esta aproximação?

$$\left.\frac{\partial S}{\partial \beta}\right|_{\beta_1} \cong \frac{S(\beta)-S(\beta_1)}{\beta-\beta_1} - \frac{1}{2}\left.\frac{\partial^2 S}{\partial \beta^2}\right|_{\beta_1}(\beta-\beta_1)$$



A equação aproxima a tangente no ponto *A* pela inclinação da reta unindo *DA* menos $\frac{1}{2}\left.\frac{\partial^2 S}{\partial \beta^2}\right|_{\beta_1}(\beta-\beta_1)$. A magnitude do termo que é subtraído de *DA* depende de quão longe β está de β_1 e da taxa na qual a inclinação da função S(β) está mudando.

Para achar o valor de β que minimiza $S(\beta)$, dado β_1 inicial, diferenciamos

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

com relação à β.

$$\frac{\partial S}{\partial \beta} \cong \left.\frac{\partial S}{\partial \beta}\right|_{\beta_1} + \left.\frac{\partial^2 S}{\partial \beta^2}\right|_{\beta_1} (\beta - \beta_1)$$

Igualando a zero e resolvendo para β obtém-se

$$\left.\frac{\partial S}{\partial \beta}\right|_{\beta_1} + \left.\frac{\partial^2 S}{\partial \beta^2}\right|_{\beta_1} \beta - \left.\frac{\partial^2 S}{\partial \beta^2}\right|_{\beta_1} \beta_1 = 0$$

$$\beta_2 = \beta_1 - \left(\left. \frac{\partial^2 S}{\partial \beta^2} \right|_{\beta_1} \right)^{-1} \left. \frac{\partial S}{\partial \beta} \right|_{\beta_1}$$

Continuando o processo obtém-se,

$$\beta_{n+1} = \beta_n - \left(\left. \frac{\partial^2 S}{\partial \beta^2} \right|_{\beta_n} \right)^{-1} \left. \frac{\partial S}{\partial \beta} \right|_{\beta_n}$$

O processo converge se $\beta_{n+1}=\beta_n$, então $\partial S/\partial\beta$ deve ser zero, que é a condição de primeira ordem.

Relação entre o algorítmo de Gauss-Newton e Newton-Raphson

Algorítmo de Gauss-Newton:

$$\beta_{n+1} = \beta_n - \frac{1}{2}\left[Z(\beta_n)' Z(\beta_n) \right]^{-1} \left. \frac{\partial S}{\partial \beta} \right|_{\beta_n}$$

$$\frac{\partial S}{\partial \beta} = 2\sum\left(Y - f\left(X_t, \beta\right)\right)\left(-\frac{\partial f\left(X_t, \beta\right)}{\partial \beta}\right)$$

$$\frac{\partial^2 S}{\partial \beta^2} = 2\sum\left[\left(\frac{\partial f\left(X_t, \beta\right)}{\partial \beta}\right)^2 - \left(Y - f\left(X_t, \beta\right)\right)\frac{\partial^2 f\left(X_t, \beta\right)}{\partial \beta^2}\right]$$

ou na forma matricial,

$$\frac{\partial^2 S}{\partial \beta^2} = 2Z(\beta)'Z(\beta) - 2\sum \left(Y - f\left(X_t, \beta\right)\right)\frac{\partial^2 f\left(X_t, \beta\right)}{\partial \beta^2}$$

então, a diferença entre o estimador de Gauss-Newton e Newton-Raphson é o termo $-2\sum\left(Y - f\left(X_t, \beta\right)\right)\frac{\partial^2 f\left(X_t, \beta\right)}{\partial \beta^2}$, o qual tem esperança zero, uma vez que a E(Y)=f(X,β).

A matriz de covariância é:

$$\hat{\sum}_b = 2\hat{\sigma}^2\left[\frac{\partial^2 S}{\partial \beta^2}\right]^{-1}\Bigg|_b$$

Como $E(Y) = f(X,\beta)$ então a

$$E\left(\frac{\partial^2 S}{\partial \beta^2}\right) = 2\left[Z(\beta)' Z(\beta)\right]$$

$$E\left(\hat{\Sigma}_b\right) = 2\hat{\sigma}^2\left[2Z(\beta)' Z(\beta)\right]^{-1} = \hat{\sigma}^2\left[Z(\beta)' Z(\beta)\right]^{-1}$$

Exemplo:

Queremos achar β que minimize

$$S(\beta) = \sum (Y - \beta X_1 - \beta^2 X_2)^2$$

$$\frac{\partial S}{\partial \beta} = -2\sum \left(X_{1t} + 2\beta X_{2t}\right)\left(Y_t - \beta X_{1t} - \beta^2 X_{2t}\right)$$

$$\frac{\partial^2 S}{\partial \beta^2} = -2\sum \left[-(X_1 + 2\beta X_2)(X_1 + 2\beta X_2) + (Y - \beta X_1 - \beta^2 X_2)(2X_2)\right]$$

$$\frac{\partial^2 S}{\partial \beta^2} = 2\sum \left[(X_1 + 2\beta X_2)^2 - (Y - \beta X_1 - \beta^2 X_2)(2X_2)\right]$$

$$\beta_2 = \beta_1 + \frac{2\sum(Y - \beta_1 X_1 - \beta_1^2 X_2)(X_1 + 2\beta_1 X_2)}{2\sum[(X_1 + 2\beta_1 X_2)^2 - (Y - \beta_1 X_1 - \beta_1^2 X_2)(2X_2)]}$$

sendo β_1=1

$$\beta_2 = 1 + \frac{\sum(YX_1 + 2X_2 Y - X_1^2 - 2X_1X_2 - X_1X_2 - 2X_2^2)}{\sum[(X_1^2 + 2\cdot 2X_1X_2 + 4X_2^2 - 2YX_2 + 2X_1X_2 + 2X_2^2]}$$

$$\beta_2 = 1 + \frac{100 + 2\cdot 96 - 30 - 3\cdot 10 - 2\cdot 19}{30 + 4\cdot 10 + 4\cdot 19 - 2\cdot 96 + 2\cdot 10 + 2\cdot 19}$$

Generalizando o modelo para K parâmetros.

$$E(\varepsilon) = 0$$

$$Y = f(X, \beta) + \varepsilon$$

$$E(\varepsilon\varepsilon') = \sigma^2 I$$

queremos encontrar o valor de β que minimiza,

$$S(\beta) = [Y - f(X,\beta)]'[Y - f(X,\beta)] = \sum_{t=1}^{T} [Y_t - f(X_t, \beta)]^2$$

Iniciamos substituindo $S(\beta)$ pela aproximação de Taylor de 2ª. ordem.

$$S(\beta) \cong S(\beta_1) + \left.\frac{\partial S}{\partial \beta'}\right|_{\beta_1} (\beta - \beta_1) + \frac{1}{2}(\beta - \beta_1) \left.\frac{\partial^2 S}{\partial\beta\partial\beta'}\right|_{\beta_1} (\beta - \beta_1)' = Q(\beta)$$

$$\frac{\partial Q(\beta)}{\partial \beta} = \left.\frac{\partial S}{\partial \beta}\right|_{\beta_1} + \left.\frac{\partial^2 S}{\partial \beta \partial \beta'}\right|_{\beta_1} (\beta - \beta_1) = 0$$

$$\left.\frac{\partial S}{\partial \beta}\right|_{\beta_1} + \left.\frac{\partial^2 S}{\partial \beta \partial \beta'}\right|_{\beta_1} \beta - \left.\frac{\partial^2 S}{\partial \beta \partial \beta'}\right|_{\beta_1} \beta_1 = 0$$

$$\left.\frac{\partial^2 S}{\partial \beta \partial \beta'}\right|_{\beta_1} \beta = \left.\frac{\partial^2 S}{\partial \beta \partial \beta'}\right|_{\beta_1} \beta_1 - \left.\frac{\partial S}{\partial \beta}\right|_{\beta_1} = 0$$

$$\beta_2 = \beta_1 - \left[\left.\frac{\partial^2 S}{\partial \beta \partial \beta'}\right|_{\beta_1} \right]^{-1} \left.\frac{\partial S}{\partial \beta}\right|_{\beta_1}$$

Generalizando,

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

onde

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

e

$$H_{K\times K} = \frac{\partial^2 S}{\partial\beta\partial\beta'} = \left. \begin{bmatrix} \frac{\partial^2 S}{\partial\beta_1^2} & \cdots & \frac{\partial^2 S}{\partial\beta_1\partial\beta_K} \\ \vdots & & \\ \frac{\partial^2 S}{\partial\beta_K\partial\beta_1} & & \frac{\partial^2 S}{\partial\beta_K^2} \end{bmatrix} \right|_{\beta_n}$$