Lecture 9
NLLS

1

M-Estimation

• An extremum estimator is one obtained as the optimizer of a criterion function, q(**z**, **b**).

Examples:

OLS: $\mathbf{b}$ = arg max $\{-\sum_{i=1}^{T} e_i^{\ 2} = -\boldsymbol{e'e}\ /T\}$

MLE: $\mathbf{b}_{\text{MLE}}$ = arg max $\{\ \ln L = \sum_{i=1}^{T} lnf(\boldsymbol{x}_i, y_i, \mathbf{b})\ \}$

GMM: $\mathbf{b}_{\text{GMM}}$ = arg max $\{-\ g(\boldsymbol{x}_i, y_i, \mathbf{b})'\ \mathbf{W}\ g(\boldsymbol{x}_i, y_i, \mathbf{b})\}$

• There are two classes of extremum estimators:

- M-estimators: The objective function is a sample average or a sum.

- Minimum distance estimators: The objective function is a measure of a *distance*.

• "M" stands for a maximum or minimum estimators –Huber (1967).

M-Estimation

- The objective function is a sample average or a sum.
- We want to minimize a population (first) moment:

$$\min_{\mathbf{b}} \text{E}[q(\boldsymbol{z}, \beta)]$$

– Using the LLN, we move from the population first moment to the sample average:

$$\sum_{i=1}^{T} q(\mathbf{z}_i, \mathbf{b})/T \xrightarrow{p} \text{E}[q(\mathbf{z}, \beta)]$$

– We want to obtain: $\mathbf{b}$ = argmin $\sum_{i=1}^{T} q(\mathbf{z}_i, \mathbf{b})$ (or divided by T)

– In general, we solve the f.o.c. (or zero-score condition):

$$\text{Zero-Score:} \quad \sum_{i=1}^{T} \frac{\partial q(\mathbf{z}_i, \mathbf{b})}{\partial \mathbf{b}'} = \mathbf{0}$$

– To check the s.o.c., we define the (pd) Hessian:

$$\mathbf{H} = \sum_{i=1}^{T} \frac{\partial^2 q(\mathbf{z}^i, \mathbf{b})}{\partial \mathbf{b}\, \partial \mathbf{b}'}$$

M-Estimation

- If $\mathbf{s}(\boldsymbol{z}, \mathbf{b}) = \frac{\partial q(\mathbf{z}_i, \mathbf{b})}{\partial \mathbf{b}'}$ exists (almost everywhere), we solve

$$\sum_{i=1}^{T} s(\mathbf{z}_i, \mathbf{b})/T = 0 \qquad (*)$$

- If, in addition, $\text{E}_X[s(\mathbf{z}_i, \mathbf{b})] = \partial/\partial\mathbf{b}'\, \text{E}_X[q(\mathbf{z}, \beta)]$ –i.e., differentiation and integration are exchangeable–, then

$$\text{E}_X[\frac{\partial q(\mathbf{z}_i, \beta)}{\partial \beta'}] = \mathbf{0}.$$

- Under these assumptions the M-estimator is said to be of *ψ-type* (ψ= $\mathbf{s}(\mathbf{z}, \mathbf{b})$ = score). Often, $\mathbf{b}_M$ is taken to be the solution of (*) without checking whether it is indeed a minimum).

- Otherwise, the M-estimator is of *ρ-type*. ($\rho = q(\mathbf{z}_i, \mathbf{b})$).

M-Estimation: LS & ML

- **Least Squares**
  - DGP: $y_i = f(x_{1,i}, x_{2,i}, \ldots, x_{k,i}; \beta) + \varepsilon_i$, $\mathbf{z} = [\boldsymbol{y}, x]$
  - $q(\boldsymbol{z}, \beta) = S(\boldsymbol{\beta}) = \sum_{i=1}^{T} \varepsilon_i^{\,2} = \sum_{i=1}^{T} (y_i - f(x_i, \beta))^2$
  - Now, we move from population to sample moments
  - $q(\boldsymbol{z}, \mathbf{b}) = S(\mathbf{b}) = \sum_{i=1}^{T} e_i^{\,2} = \sum_{i=1}^{T} (y_i - f(x_i, \mathbf{b})^2$
  - $\mathbf{b}_{NLLS}$ = argmin S(**b**)
- **Maximum Likelihood**
  - Let $f(x_i, \beta)$ be the pdf of the data.
  - $L(\boldsymbol{x}, \beta) = \prod_{i=1}^{T} f(x_i, \beta)$
  - $\ln L(\boldsymbol{x}, \beta) = \sum_{i=1}^{T} \ln f(x_i, \beta)$
  - Now, we move from population to sample moments
  - $q(\boldsymbol{z}, \mathbf{b}) = -\ln L(\boldsymbol{x}, \mathbf{b})$
  - $\mathbf{b}_{MLE}$ = argmin $\{ -\ln L(\boldsymbol{x}, \mathbf{b}) \}$

M-Estimation: Minimum L_p-estimators

• Minimum L_p-estimators

- $q(\boldsymbol{z}, \beta) = (1/p)\,|\boldsymbol{x} - \beta|^p$ for $1 \le p \le 2$
- $\boldsymbol{s}(\boldsymbol{z}, \beta) = |\boldsymbol{x} - \beta|^{p-1}$ $\quad \boldsymbol{x} - \beta < 0$
$\qquad = -|\boldsymbol{x} - \beta|^{p-1}$ $\quad \boldsymbol{x} - \beta > 0$

• Special cases:

– $p = 2$: We get the sample mean (LS estimator for β).

$$\boldsymbol{s}(\boldsymbol{z}, \beta) = \sum_{i=1}^{T} (x_i - \mathbf{b}_M) = 0 \quad \Rightarrow \mathbf{b}_M = \sum_{i=1}^{T} x_i / T$$

– $p = 1$: We get the sample median as the estimator with the least absolute deviation (LAD) for the median β. (There is no unique solution if T is even.)

Note: Unlike LS, LAD does not have an analytical solving method. Numerical optimization is not feasible. Linear programming is used.

The Score Vector

• Let $\boldsymbol{X}$ = {X_1; X_2;... }be *i.i.d.*

• If $\boldsymbol{s}(\boldsymbol{z}, \boldsymbol{b}) = \partial q(\boldsymbol{z}, \beta)/\partial \mathbf{b}'$ exists, we solve

$$\sum_{i=1}^{T} \boldsymbol{s}(\boldsymbol{z}_i, \boldsymbol{b}_M)/T = 0 \qquad (\boldsymbol{s}(\boldsymbol{z}_i, \boldsymbol{b}) \text{ is a } k\text{x1 vector}).$$

– $E[\boldsymbol{s}(\boldsymbol{z}, \boldsymbol{b}_0)] = E[\partial\, q(\boldsymbol{z}, \mathbf{b})/\partial \mathbf{b}'] = \mathbf{0}$

– Using the LLN: $\sum_{i=1}^{T} \boldsymbol{s}(\boldsymbol{z}_i, \boldsymbol{b}_M)/T \xrightarrow{p} E[\boldsymbol{s}(\boldsymbol{z}, \boldsymbol{b}_0)] = \mathbf{0}$

– $\boldsymbol{V} = \text{Var}[\boldsymbol{s}(\boldsymbol{z}, \boldsymbol{b}_0)] = E[\boldsymbol{s}(\boldsymbol{z}, \boldsymbol{b}) * \boldsymbol{s}(\boldsymbol{z}, \boldsymbol{b})']$ ($\boldsymbol{V}$ is a kxk matrix).

$= E[(\partial q(\boldsymbol{z}, \mathbf{b})/\partial \mathbf{b}') * (\partial q(\boldsymbol{z}, \mathbf{b})/\partial \mathbf{b}')']$

– Using the LLN: $\sum_{i=1}^{T} [\boldsymbol{s}(\boldsymbol{z}_i, \boldsymbol{b}_M)\, \boldsymbol{s}(\boldsymbol{z}_i, \boldsymbol{b}_M)']/T \xrightarrow{p} \text{Var}[\boldsymbol{s}(\boldsymbol{z}, \boldsymbol{b}_0)]$

– Using the Lindeberg-Levy CLT: $\sum_{i=1}^{T} \boldsymbol{s}(\boldsymbol{z}_i, b)/\sqrt{T} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{V})$

<u>Note</u>: We have already shown these results for the ML case.

The Hessian Matrix

• $\boldsymbol{H}(\mathbf{z}, \mathbf{b}) = E[\partial \mathbf{s}(\mathbf{z}, \mathbf{b})/\partial \mathbf{b}] = E[\partial^2 q(\mathbf{z}; \mathbf{b})/\partial \mathbf{b} \partial \mathbf{b}']$

- Using the LLN: $\sum_{i=1}^{T} [\partial \boldsymbol{s}(\boldsymbol{z}_i, \boldsymbol{b}_M)/\partial \mathbf{b}]/T \xrightarrow{p} \boldsymbol{H}(\mathbf{z}, \mathbf{b}_0)$

• In general, the Information (Matrix) Equality does not hold. That is, $\boldsymbol{H} \neq \boldsymbol{V}$. The equality only holds if the model is correctly specified.

The Asymptotic Theory

• We have all the tools to derive the asymptotic distribution of $\boldsymbol{b}_M$.

Recall the Mean Value Theorem:

$$f(x) = f(a) + f'(b)\ (x - a) \qquad a < b < x$$

Apply MVT to the score, with $\boldsymbol{b}_0 < \mathbf{b}* < \boldsymbol{b}_M$:

$$\sum_{i=1}^{T} \boldsymbol{s}(\boldsymbol{z}_i, \boldsymbol{b}_M) = \sum_{i=1}^{T} \boldsymbol{s}(\boldsymbol{z}_i, \boldsymbol{b}_0) + \sum_{i=1}^{T} \boldsymbol{H}(\boldsymbol{z}_i, \mathbf{b}*)\ (\mathbf{b_M} - \mathbf{b_0})$$

$$0 = \sum_{i=1}^{T} \boldsymbol{s}(\boldsymbol{z}_i, \boldsymbol{b}_0) + \sum_{i=1}^{T} \boldsymbol{H}(\boldsymbol{z}_i, \mathbf{b}*)\ (\mathbf{b_M} - \mathbf{b_0})$$

$$\Rightarrow \quad (\boldsymbol{b}_M - \mathbf{b_0}) = [\sum_{i=1}^{T} \boldsymbol{H}(\boldsymbol{z}_i, \mathbf{b}*)]^{-1} \sum_{i=1}^{T} \boldsymbol{s}(\boldsymbol{z}_i, \boldsymbol{b}_0)$$

$$\Rightarrow \sqrt{T}\ (\boldsymbol{b}_M - \mathbf{b_0}) = [\sum_{i=1}^{T} \boldsymbol{H}(\boldsymbol{z}_i, \mathbf{b}*)]^{-1} \sum_{i=1}^{T} \boldsymbol{s}(\boldsymbol{z}_i, \boldsymbol{b}_0)/\sqrt{T}$$

The asymptotic distribution of $\boldsymbol{b}_M$ is driven by $\sum_{i=1}^{T} \boldsymbol{s}(\boldsymbol{z}_i, \boldsymbol{b}_0)/\sqrt{T}$

The Asymptotic Theory

• **Theorem**: Consistency of M-estimators

Let $\{\boldsymbol{X} = X_1; X_2;... \}$ be *i.i.d.* and assume

(1) $\mathbf{b} \in \mathbf{B}$, where $\mathbf{B}$ is compact. (*"compact"*)

(2) $[\sum_i q(\mathbf{X}_i; \mathbf{b})/T] \xrightarrow{p} g(\mathbf{b})$ uniformly in $\mathbf{b}$ for some continuous function g: $\mathbf{B} \rightarrow R$ (*"continuity"*)

(3) $g(\mathbf{b})$ has a unique global minimum at $\mathbf{b_0}$. (*"identification"*)

Then, $\boldsymbol{b}_M \xrightarrow{p} \mathbf{b_0}$

<u>Remark</u>: a) Since $\boldsymbol{X}$ are *i.i.d.* by the LLN (without uniformity) it must hold $g(\mathbf{b}) = E_X[q(\boldsymbol{X}; \mathbf{b})]$, thus $E_X[q(\mathbf{z}, \mathbf{b_0})] = \min_{\mathbf{b} \in \mathbf{B}} E_X[q(\mathbf{z}; \beta)]$.

b) If $\mathbf{B}$ is not compact, find a compact subset $\mathbf{B_0}$, with $\mathbf{b_0} \in \mathbf{B_0}$ and $P[\mathbf{b_M} \in \mathbf{B_0}] \rightarrow 1$.

The Asymptotic Theory

Theorem: Asymptotic Normality of M-estimators

Assumptions:

(1) $\boldsymbol{b}_M \xrightarrow{p} \boldsymbol{b}_0$ for some $\boldsymbol{b}_0 \in \mathbf{B}$.

(2) $\boldsymbol{b}_M$ is of *ψ-type* and **s** is continuously (for almost all x) differentiable w.r.t. **b**.

(3) $\sum_{i=1}^{T}[\partial \boldsymbol{s}(\boldsymbol{z}_i, \boldsymbol{b})/\partial \mathbf{b}]/T|_{\mathbf{b}=\mathbf{b}*} \xrightarrow{p} \boldsymbol{H}(\mathbf{z}, \boldsymbol{b}_0)$ for $\mathbf{b}* \xrightarrow{p} \boldsymbol{b}_0$

(4) $\sum_{i=1}^{T} \boldsymbol{s}(\boldsymbol{z}_i, \boldsymbol{b})/\sqrt{T} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{V}_0)$ $\qquad \boldsymbol{V}_0 = \text{Var}[\mathbf{s}(\mathbf{z}, \boldsymbol{b}_0)] < \infty$

Then, $\sqrt{T}\,(\boldsymbol{b}_M - \mathbf{b}_0) = [\sum_{i=1}^{T} \boldsymbol{H}(\boldsymbol{z}_i, \mathbf{b}*)]^{-1} \sum_{i=1}^{T} \boldsymbol{s}(\boldsymbol{z}_i, \boldsymbol{b}_0)$

$\Rightarrow \sqrt{T}\,(\boldsymbol{b}_M - \boldsymbol{b}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{H}_0^{-1} \boldsymbol{V}_0 \boldsymbol{H}_0^{-1})$

- $\boldsymbol{V} = \text{E}[\mathbf{s}(\mathbf{z},\mathbf{b})\, \mathbf{s}(\mathbf{z},\mathbf{b})'] = \text{E}[(\partial \text{q}(\mathbf{z},\mathbf{b})/\partial \mathbf{b})' (\partial \text{q}(\mathbf{z},\mathbf{b})/\partial \mathbf{b})]$

- $\boldsymbol{H} = \partial \mathbf{s}(\mathbf{z},\mathbf{b})/\partial \mathbf{b} = \text{E}[\partial^2 \text{q}(\mathbf{z},\mathbf{b})/\partial \mathbf{b} \partial \mathbf{b}']$

Asymptotic Normality

- Summary
  - $\boldsymbol{b}_M \xrightarrow{p} \boldsymbol{b}_0$
  - $\boldsymbol{b}_M \xrightarrow{a} N(\boldsymbol{b}_0, \text{Var}[\boldsymbol{b}_0])$
  - $\text{Var}[\boldsymbol{b}_M] = (1/T)\, \boldsymbol{H}_0^{-1} \boldsymbol{V}_0 \boldsymbol{H}_0^{-1}$
  - If the model is correctly specified: -**H** = **V**.
    Then, $\text{Var}[\mathbf{b}] = \boldsymbol{V}_0$

  – **H** and **V** are evaluated at $\mathbf{b}_0$:

  - $\mathbf{H} = \sum_i [\partial^2 \text{q}(\mathbf{z}_i; \mathbf{b})/\partial \mathbf{b} \partial \mathbf{b}']$

  - $\mathbf{V} = \sum_i [\partial \text{q}(\mathbf{z}_i; \mathbf{b})/\partial \mathbf{b}][\partial \text{q}(\mathbf{z}_i; \mathbf{b})/\partial \mathbf{b}']$

M-Estimation: Example

• DGP: $\mathbf{y} = f(\mathbf{x}_i;\beta) + \boldsymbol{\varepsilon} = \exp(\mathbf{x}\beta) + \boldsymbol{\varepsilon}$,

- Objective function:

$$q(\mathbf{X};\ \beta) = ½\ \boldsymbol{\varepsilon}'\ \boldsymbol{\varepsilon} = ½\ [\mathbf{y} - \exp(\mathbf{X}\beta)]'\ [\mathbf{y} - \exp(\mathbf{X}\beta)]$$

Let $\mathbf{G} = [g_i]$, where $g_i = \partial f_i\,(\mathbf{x};\ \beta)/\partial\beta_{\mathbf{k}} = \exp(\mathbf{x}_i'\beta)\ \mathbf{x}_{\mathbf{k}i}$

- Score: $\mathbf{s}(\mathbf{z},\ \beta) = \partial q(\mathbf{z};\ \beta)/\partial\beta = \boldsymbol{\varepsilon}'\ \partial f(\mathbf{x}_i;\ \beta)/\partial\beta = \boldsymbol{\varepsilon}'\ \mathbf{G}$

$$= -\ [\mathbf{y} - \exp(\mathbf{X}\beta)]'\ \mathbf{G} = -\ \mathbf{y}'\mathbf{G} + \exp(\mathbf{X}\beta)'\ \mathbf{G}$$

- $\boldsymbol{V} = \text{Var}[\mathbf{s}(\mathbf{z},\beta)] = \text{E}[\mathbf{G}'\boldsymbol{\varepsilon}\ \boldsymbol{\varepsilon}'\mathbf{G}]$

- $\boldsymbol{H} = \text{E}[\partial^2 q(\mathbf{z},\ \beta)/\partial\beta\partial\beta'] = \text{E}[\partial f(\mathbf{x}_i;\beta)/\partial\beta'\ \partial f(\mathbf{x}_i;\ \beta)/\partial\beta' -$

$$- \partial^2 f(\mathbf{x}_i;\ \beta)/\partial\beta\partial\beta'\ \boldsymbol{\varepsilon}] = \text{E}[\mathbf{G}'\mathbf{G} - \partial\mathbf{G}/\partial\beta\ \boldsymbol{\varepsilon}]$$

- $\text{Var}[\mathbf{b_M}] = (1/T)\ \boldsymbol{H}_\mathbf{0}^{-1}\ \boldsymbol{V}_\mathbf{0}\ \boldsymbol{H}_\mathbf{0}^{-1}$

M-Estimation: Example

• $\text{Var}[\ \boldsymbol{b}_M] = (1/T)\ \boldsymbol{H}_\mathbf{0}^{-1}\ \boldsymbol{V}_\mathbf{0}\ \boldsymbol{H}_\mathbf{0}^{-1}$

• We approximate ("estimate")

$$\text{Var}[\boldsymbol{b}_M] = (1/T)\ \{\textstyle\sum_i [\partial\mathbf{s}(\mathbf{z}_i,\ \boldsymbol{b}_M)/\partial\mathbf{b}]\}^{-1}\ [\textstyle\sum_i \mathbf{s}(\mathbf{z}_i,\ \boldsymbol{b}_M)\ \mathbf{s}(\mathbf{z}_i,\ \boldsymbol{b}_M)']\ \{\textstyle\sum_i [\partial\mathbf{s}(\mathbf{z}_i,\ \boldsymbol{b}_M)/\partial\mathbf{b}]\}^{-1}$$

$$\mathbf{s}(\mathbf{z}_i,\ \boldsymbol{b}_M) = -\ [\exp(\mathbf{x}_i'\ \boldsymbol{b}_M)\mathbf{x}_i]'\ [y_i - \exp(\mathbf{x}_i'\ \boldsymbol{b}_M)] = -\ \mathbf{x}_i'\ \exp(\mathbf{x}_i'\ \boldsymbol{b}_M)'\ \mathbf{e}_i$$

Two-Step M-Estimation

• Sometimes, nonlinear models depend not only on our parameter of interest $\boldsymbol{\beta}$, but nuisance parameters or unobserved variables in some way. It is common to estimate $\boldsymbol{\beta}$ using a "two-step" procedure:

1st-stage: $\mathbf{y}_2 = g(\mathbf{w}; \boldsymbol{\gamma}) + \boldsymbol{\nu}$ $\Rightarrow$ we estimate $\boldsymbol{\gamma}$, say $\mathbf{c}$

2nd-stage $\mathbf{y} = f(\mathbf{x}; \boldsymbol{\beta},\mathbf{c}) + \boldsymbol{\varepsilon}$ $\Rightarrow$ we estimate $\boldsymbol{\beta}$, given $\mathbf{c}$.

• The objective function: $\min_{\beta} \{\sum_i q(\mathbf{x}; \boldsymbol{\beta},\mathbf{c}) = \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}\}$

• **Examples**:

(i) DHW Test for endogeneity

(ii) Weighted NLLS: $\min_{\beta} \{\sum_i [\mathbf{y} - f(\mathbf{x}; \boldsymbol{\beta})]^2 / g(\mathbf{z}; \mathbf{c})$

(iii) Selection Bias Model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta}\, \hat{h} + \boldsymbol{\varepsilon}$ $\quad \hat{h} = G(\mathbf{z}, \mathbf{c})$.

Two-Step M-Estimation

• Properties --Pagan (1984, 1986), *generated regressors*:

- Consistency. We need to apply a uniform weak LLN.
- Asymptotic normality: We need to apply CLT.

• Two interesting results:

- The 2S estimator can be consistent even in some cases where $g(\mathbf{z};\boldsymbol{\gamma})$ is not correctly specified –i.e., situations where $\mathbf{c}$ may be inconsistent.

- The S.E. –i.e., Var[$\mathbf{b}_{2S}$]– needs to be adjusted by the 1st stage estimation, in most cases.

Two-Step M-Estimation

• Recall $\sqrt{T}\,(\boldsymbol{b}_M - \mathbf{b}_0) = \boldsymbol{H}_0^{-1}[-\sum_i \mathbf{s}(\mathbf{z}_i;\, \mathbf{b}_0,\mathbf{c})/\sqrt{T}] + o(1)$ (*)

The question is weather the following equation holds:

$$\sum_i \mathbf{s}(\mathbf{z}_i;\, \mathbf{b}_0,\mathbf{c})/\sqrt{T} = \sum_i \mathbf{s}(\mathbf{z}_i;\, \mathbf{b}_0,\mathbf{c}_0)/\sqrt{T} + o(1) \qquad (**)$$

where $\mathbf{c}_0$ is the true value of $\boldsymbol{\gamma}$.

If this equality holds, $\mathbf{b}_M$ would be consistent.

• Let's do a 1st order Taylor expansion:

$$\sum_i \mathbf{s}(\mathbf{z}_i;\, \mathbf{b}_0,\mathbf{c})/\sqrt{T} \approx \sum_i \mathbf{s}(\mathbf{z}_i;\, \mathbf{b}_0,\mathbf{c}_0)/\sqrt{T} + \boldsymbol{F}_0\,(\mathbf{c} - \mathbf{c}_0)/\sqrt{T} \qquad (***)$$

where $\boldsymbol{F}_0 = \partial \mathbf{s}(\mathbf{z};\, \mathbf{b}_0,\mathbf{c})/\partial \boldsymbol{\gamma}$

Note: If $\mathbf{c} = \mathbf{c}_0$ or $\boldsymbol{F}_0 = 0$, then (**) holds.

Two-Step M-Estimation

• We can also write

$$\sqrt{T}\,(\mathbf{c} - \mathbf{c}_0) = \boldsymbol{H}_{\mathbf{c}0}^{-1}[-\sum_i \mathbf{s}(\mathbf{w}_i,\, \mathbf{c})/\sqrt{T}] + o(1)$$
$$= \sum_i \mathbf{h}(\mathbf{w}_i,\, \mathbf{c})/\sqrt{T} + o(1)$$

• Then, substituting back in (***) and then in (*), we have

$$\sqrt{T}\,(\boldsymbol{b}_M - \mathbf{b}_0) = \boldsymbol{H}_0^{-1}[-\sum_i \mathbf{r}(\mathbf{z}_i;\, \mathbf{b}_0,\mathbf{c}_0)/\sqrt{T}] + o(1), \qquad (****)$$

where $\mathbf{r}(\mathbf{z}_i;\, \mathbf{b}_0,\mathbf{c}_0) = \mathbf{s}(\mathbf{z}_i;\, \mathbf{b}_0,\mathbf{c}_0) + \boldsymbol{F}_0\, \mathbf{h}(\mathbf{w}_i,\, \mathbf{c}_0)$

Note: Difference between (*) and (****): $\mathbf{r}(\mathbf{z}_i,\, \mathbf{b}_0,\mathbf{c}_0)$ replaces $\mathbf{s}(\mathbf{z}_i,\mathbf{b}_0,\mathbf{c})$. The second term in $\mathbf{r}(\mathbf{z}_i,\, \mathbf{b}_0,\mathbf{c}_0)$ reflects the 1st-stage adjustment.

• $\text{Var}[\boldsymbol{b}_M] = (1/T)\, \boldsymbol{H}_0^{-1}\, \mathbf{Var}[\mathbf{r}(\mathbf{z}_i;\, \mathbf{b}_0,\mathbf{c}_0)]\, \boldsymbol{H}_0^{-1}$

Applications

- Heteroscedastity Autocorrelation Consistent (HAC) Variance-Covariance Matrix
  - Non-spherical disturbances in NLLS

- Quasi Maximum Likelihood (QML)
  - Misspecified density assumption in ML
  - Information Equality may not hold

Special case of M-estimation: NL Regression

• We start with a regression model: $y_i = f(\boldsymbol{x}_i, \boldsymbol{\beta}) + \varepsilon_i$

• Q: What makes a regression model *nonlinear*?

• Recall that OLS can be applied to nonlinear functional forms. But, for OLS to work, we need *intrinsic linearity* –i.e., the model linear in the parameters.

Example: A nonlinear functional form, but intrinsic linear:

$$y_i = \exp(\beta_1) + \beta_2 * x_i + \beta_3 * x_i^2 + \varepsilon_i$$

Example: A non intrinsic linear model:

$$y_i = f(\boldsymbol{x}_i, \boldsymbol{\beta}^0) + \varepsilon_i = \beta_0 + \beta_1 x_i^{\beta_2} + \varepsilon_i$$

Nonlinear Least Squares

• Least squares:

$$\text{Min}_{\beta} \{ S(\beta) = ½ \sum_{i=1}^{T} [y_i - f(\boldsymbol{x}_i, \beta)]^2 = ½ \sum_{i=1}^{T} \varepsilon_i^{\,2} \}$$

F.o.c.:

$$\partial\{½\sum_{i=1}^{T} [y_i - f(\boldsymbol{x}_i, \beta)]^2\}/\partial\beta$$
$$= ½\, 2 \sum_{i=1}^{T} [y_i - f(\boldsymbol{x}_i, \beta)]^2 \,\partial f(\boldsymbol{x}_i, \beta)/\partial\beta = -\sum_{i=1}^{T} e_i\, \boldsymbol{x}_i^{\mathbf{0}}$$
$$\Rightarrow \quad -\sum_{i=1}^{T} e_i\, \boldsymbol{x}_i^{\mathbf{0}} = \mathbf{0}, \qquad \text{we solve for } \mathbf{b}_{NLLS}.$$

In general, there is no explicit solution, like in the OLS case:

$$\mathbf{b} = g(\mathbf{X}, \mathbf{y}) = (\mathbf{X'X})^{-1}\mathbf{X'}\,\boldsymbol{y}$$

• In this case, we have a *nonlinear* model: the f.o.c. cannot be solved explicitly for $\mathbf{b}_{NLLS}$. That is, the nonlinearity of the f.o.c. defines a nonlinear model.

Nonlinear Least Squares: Example

• Q: How to solve this kind of set of equations?

Example: $\text{Min}_{\beta} \{ S(\beta) = ½ \sum_{i=1}^{T} [y_i - f(\boldsymbol{x}_i, \beta)]^2 = ½ \sum_{i=1}^{T} \varepsilon_i^{\,2} \}$

$$y_i = f(\boldsymbol{x}_i, \beta) + \varepsilon_i = \beta_0 + \beta_1\, x_i^{\beta_2} + \varepsilon_i.$$

f.o.c.:

$$\partial[\,½\sum_{i=1}^{T} e_i^{\,2}]/\partial\beta_0 = \Sigma_i\, (-1)\, (y_i - \beta_0 + \beta_1\, x_i^{\beta 2})\;\; 1 = 0$$
$$\partial[\,½\sum_{i=1}^{T} e_i^{\,2}]/\partial\beta_1 = \Sigma_i\, (-1)\, (y_i - \beta_0 + \beta_1\, x_i^{\beta 2})\; x_i^{\beta 2} = 0$$
$$\partial[\,½\sum_{i=1}^{T} e_i^{\,2}]/\partial\beta_2 = \Sigma_i\, (-1)\, (y_i - \beta_0 + \beta_1\, x_i^{\beta 2})\; \beta_1 x_i^{\beta 2} \ln(x_i) = 0$$

• Nonlinear equations require a nonlinear solution. This defines a nonlinear regression model: the f.o.c. are *not* linear in $\boldsymbol{\beta}$.

Note: If $\beta_2 = 1$, we have a linear model.

Nonlinear Least Squares: Example

Example: Min $_\beta$ $S(\boldsymbol{\beta}) = \{ ½ \sum_{i=1}^{T} [y_i - (\beta_0 + \beta_1 x_i^{\beta_2} + \varepsilon_i)]^2 \}$

• From the f.o.c., we cannot solve for $\boldsymbol{\beta}$ explicitly. But, using some steps, we can still minimize RSS to obtain estimates of $\boldsymbol{\beta}$.

• Nonlinear regression algorithm:

1. Start by guessing a plausible values for $\boldsymbol{\beta}$, say $\boldsymbol{\beta}^0$.
2. Calculate RSS for $\boldsymbol{\beta}^0$ $\Rightarrow$ get RSS($\boldsymbol{\beta}^0$)
3. Make small changes to $\boldsymbol{\beta}^0$ $\Rightarrow$ get $\boldsymbol{\beta}^1$.
4. Calculate RSS for $\boldsymbol{\beta}^1$ $\Rightarrow$ get RSS($\boldsymbol{\beta}^1$)
5. If RSS($\boldsymbol{\beta}^1$) < RSS($\boldsymbol{\beta}^0$) $\Rightarrow$ $\boldsymbol{\beta}^1$ becomes the new starting point.
6. Repeat steps 3-5 until you RSS($\boldsymbol{\beta}^j$) cannot be lowered. $\Rightarrow$ get $\boldsymbol{\beta}^j$.

$\Rightarrow$ $\boldsymbol{\beta}^j$ is the (nonlinear) least squares estimates.

4

NLLS: Linearization

• We start with a nonlinear model: $y_i = f(\boldsymbol{x}_i, \boldsymbol{\beta}) + \varepsilon_i$

• We expand the regression around some point, $\boldsymbol{\beta}^0$:

$$f(\boldsymbol{x}_i, \boldsymbol{\beta}) \approx f(\boldsymbol{x}_i, \boldsymbol{\beta}^0) + \sum_{j=1}^{k} [\partial f(\boldsymbol{x}_i, \boldsymbol{\beta}^0)/\partial \beta_j^0] * (\beta_j - \beta_j^0)$$
$$= f(\boldsymbol{x}_i, \boldsymbol{\beta}^0) + \sum_{j=1}^{k} \boldsymbol{x}_i^0 * (\beta_j - \beta_j^0)$$
$$= [f(\boldsymbol{x}_i, \boldsymbol{\beta}^0) - \sum_{j=1}^{k} \boldsymbol{x}_i^0 * \beta_j^0] + \sum_{j=1}^{k} \boldsymbol{x}_i^0 * \beta_j$$
$$= f_i^0 + \sum_{j=1}^{k} \boldsymbol{x}_i^0 * \beta_j = f_i^0 + \boldsymbol{x}_i^{0\prime} \boldsymbol{\beta}$$

where

$f_i^0 = f(\boldsymbol{x}_i, \boldsymbol{\beta}^0) - \boldsymbol{x}_i^{0\prime} \boldsymbol{\beta}^0$ (f_i^0 does not depend on unknowns)

Now, $f(\boldsymbol{x}_i, \boldsymbol{\beta})$ is (approximately) linear in the parameters! That is,

$$y_i = f_i^0 + \boldsymbol{x}_i^{0\prime} \boldsymbol{\beta} + \varepsilon_i^0 \qquad (\varepsilon_i^0 = \varepsilon_i + \text{linearization error } i)$$

$$\Rightarrow y_i^0 = y_i - f_i^0 = \boldsymbol{x}_i^{0\prime} \boldsymbol{\beta} + \varepsilon_i^0$$

NLLS: Linearization

• We linearized $f(\boldsymbol{x}_i, \boldsymbol{\beta})$ to get:

$$\boldsymbol{y} = \boldsymbol{f}^0 + \mathbf{X}^0 \boldsymbol{\beta} + \boldsymbol{\varepsilon}^0 \qquad (\boldsymbol{\varepsilon}^0 = \boldsymbol{\varepsilon} + \text{linearization error})$$

$$\Rightarrow \boldsymbol{y}^0 = \boldsymbol{y} - \boldsymbol{f}^0 = \mathbf{X}^0 \boldsymbol{\beta} + \boldsymbol{\varepsilon}^0$$

• Now, we can do OLS:

$$\mathbf{b}_{NLLS} = (\mathbf{X}^{0\prime} \mathbf{X}^0)^{-1} \mathbf{X}^{0\prime} \boldsymbol{y}^0$$

Note: $\mathbf{X}^0$ are called *pseudo-regressors*.

• In general, we get different $\mathbf{b}_{NLLS}$ for different $\boldsymbol{\beta}^0$. An algorithm can be used to get the *best* $\mathbf{b}_{NLLS}$.

• We will resort to *numerical optimization* to find the $\mathbf{b}_{NLLS}$.

NLLS: Linearization

• We can also compute the asymptotic covariance matrix for the NLLS estimator as usual, using the pseudo regressors and the RSS:

$$\text{Est. Var}[\mathbf{b}_{NLLS} | \mathbf{X}^0] = s^2_{NLLS} (\mathbf{X}^{0\prime} \mathbf{X}^0)^{-1}$$

$$s^2_{NLLS} = [\boldsymbol{y} - f(\boldsymbol{x}_i, \mathbf{b}_{NLLS})]' [\boldsymbol{y} - f(\boldsymbol{x}_i, \mathbf{b}_{NLLS})]/(T - k).$$

• Since the results are asymptotic, we do not need a degrees of freedom correction. However, a *df* correction is usually included.

Note: To calculate s^2_{NLLS}, we calculate the residuals from the nonlinear model, not from the linearized model (linearized regression).

NLLS: Linearization – Example

• Nonlinear model: $y_i = f(\boldsymbol{x}_i, \beta^0) + \varepsilon_i = \beta_0 + \beta_1 x_i^{\beta_2} + \varepsilon_i$

• Linearize the model to get:

$\boldsymbol{y}^0 = \boldsymbol{y} - \boldsymbol{f}^0 = \mathbf{X}^0 \beta + \varepsilon^0$, where $f_i^0 = f(\boldsymbol{x}_i, \beta^0) - \boldsymbol{x}_i^{0\prime} \beta^0$

Get $\boldsymbol{x}_i^0 = \partial f(\boldsymbol{x}_i, \beta)/\partial\beta|_{\beta=\beta 0}$

$\partial f(\boldsymbol{x}_i, \beta)/\partial\beta_0 = 1$

$\partial f(\boldsymbol{x}_i, \beta)/\partial\beta_1 = x_i^{\beta_2}$

$\partial f(\boldsymbol{x}_i, \beta)/\partial\beta_2 = \beta_1 x_i^{\beta_2} \ln(x_i)$

$f_i^0 = \beta_0^0 + \beta_1^0 x_i^{\beta_2^0} - \{\beta_0^0 + \beta_1^0 x_i^{\beta_2^0} + \beta_2^0 \beta_1^0 x_i^{\beta_2^0} \ln(x_i)\} \boldsymbol{x}^{\beta_2^0}$

$y_i^0 = \beta_0 + \beta_1 x_i^{\beta_2^0} + \beta_2 \beta_1^0 x_i^{\beta_2^0} \ln(x_i) + \varepsilon_i^0$

To get $\mathbf{b}_{NLLS}$, regress $\boldsymbol{y}^0$ on a constant, $\boldsymbol{x}^{\beta_2^0}$, and $\beta_1^0 \boldsymbol{x}^{\beta_2^0} \ln(\boldsymbol{x})$.

Gauss-Newton Algorithm

• Recall that $\mathbf{b}_{NLLS}$ depends on $\boldsymbol{\beta}^0$. That is,

$\mathbf{b}_{NLLS}(\boldsymbol{\beta}^0) = (\mathbf{X}^{0\prime}\mathbf{X}^0)^{-1}\mathbf{X}^{0\prime}\boldsymbol{y}^0$

• We use a *Gauss-Newton algorithm* to find the $\mathbf{b}_{NLLS}$. Recall GN:

$\beta_{k+1} = \beta_k + (\mathbf{J}^T\mathbf{J})^{-1}\mathbf{J}^T\boldsymbol{\varepsilon}$ – $\mathbf{J}$: Jacobian = $\delta f(x_i;\boldsymbol{\beta})/\delta\boldsymbol{\beta}$.

• Given a $\mathbf{b}_{NLLS}$ at step j, $\mathbf{b}(j)$, we find the $\mathbf{b}_{NLLS}$ for step j+1 by:

$\mathbf{b}(j+1) = \mathbf{b}(j) + [\mathbf{X}^0(j)'\mathbf{X}^0(j)]^{-1}\mathbf{X}^0(j)'\boldsymbol{e}^0(j)$

Columns of $\mathbf{X}^0(j)$ are the derivatives: $\partial f(x_i,\mathbf{b}(j))/\partial\mathbf{b}(j)'$

$\mathbf{e}^0(j) = \boldsymbol{y} - f[\mathbf{x},\mathbf{b}(j)]$

• The *update* vector is the slopes in the regression of the residuals on $\mathbf{X}^0$. The update is zero when they are orthogonal. (Just like OLS)

Box-Cox Transformation

• It's a simple transformation that allows non-linearities in the CLM.

$$y_i = f(\boldsymbol{x}_i, \beta) + \varepsilon_i = \sum_{j=1}^{k} \boldsymbol{x}_{i,j}^{(\lambda)}\, \beta_j + \varepsilon_i$$

$$\boldsymbol{x}_k^{(\lambda)} = (\boldsymbol{x}_k^{\lambda} - 1)/\lambda \qquad \lim_{\lambda\to 0}\ (\boldsymbol{x}_k^{\lambda} - 1)/\lambda = \ln \boldsymbol{x}_k$$

• For a given λ, OLS can be used. An iterative process can be used to estimate λ. OLS standard errors have to be corrected. Probably, not a very efficient method.

• NLLS or MLE will work fine.

• We can have a more general Box-Cox transformation model:

$$y_i^{(\lambda 1)} = \sum_{j=1}^{k} \boldsymbol{x}_{i,j}^{(\lambda 2)}\, \beta_j + \varepsilon_i$$

Testing non-linear restrictions

• Testing linear restrictions as before.

• Non-linear restrictions introduce slight modification to the usual tests. We want to test:

$$H_0: R(\boldsymbol{\beta}) = 0$$

where $R(\boldsymbol{\beta})$ is a non-linear function, with $\text{rank}[\partial R(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = \mathbf{G}(\boldsymbol{\beta})] = J$.

• A Wald test can be based on $\mathbf{m} = R(\mathbf{b}_{NLLS}) - \mathbf{0}$:

$$W = \mathbf{m}'(\text{Var}[\mathbf{m}|\mathbf{X}])^{-1}\mathbf{m} = R(\mathbf{b}_{NLLS})'(\text{Var}[R(\mathbf{b}_{NLLS})|\mathbf{X}])^{-1} R(\mathbf{b}_{NLLS})$$

<u>Problem</u>: We do not know the distribution of $R(\mathbf{b}_{NLLS})$, but we know the distribution of $\mathbf{b}_{NLLS}$.

<u>Solution</u>: Linearize $R(\mathbf{b}_{NLLS})$ around $\boldsymbol{\beta}$

$$R(\mathbf{b}_{NLLS}) \approx R(\boldsymbol{\beta}) + \mathbf{G}(\mathbf{b}_{NLLS})\,(\mathbf{b}_{NLLS} - \boldsymbol{\beta})$$

Testing non-linear restrictions

• Linearize $R(\mathbf{b}_{NLLS})$ around $\boldsymbol{\beta}$ (= $\mathbf{b_0}$)

$$R(\mathbf{b}_{NLLS}) \approx R(\boldsymbol{\beta}) + \mathbf{G}(\mathbf{b}_{NLLS})\,(\mathbf{b}_{NLLS} - \boldsymbol{\beta})$$

• Recall $\sqrt{T}\,(\mathbf{b_M} - \mathbf{b_0}) \xrightarrow{d} N(\mathbf{0}, \text{Var}[\mathbf{b_0}])$

where $\text{Var}[\mathbf{b_0}] = \boldsymbol{H}(\boldsymbol{\beta})^{-1}\,\boldsymbol{V}(\boldsymbol{\beta})\,\boldsymbol{H}(\boldsymbol{\beta})^{-1}$

$$\Rightarrow \sqrt{T}\,[R(\mathbf{b}_{NLLS}) - R(\boldsymbol{\beta})] \xrightarrow{d} N(\mathbf{0}, \boldsymbol{G}(\boldsymbol{\beta})\,\text{Var}[\mathbf{b_0}]\,\boldsymbol{G}(\boldsymbol{\beta})')$$

$$\Rightarrow \text{Var}[R(\mathbf{b}_{NLLS})] = (1/T)\,\boldsymbol{G}(\boldsymbol{\beta})\,\text{Var}[\mathbf{b_0}]\,\boldsymbol{G}(\boldsymbol{\beta})'$$

• Then,

$$W = T\,R(\mathbf{b}_{NLLS})'\{\boldsymbol{G}(\mathbf{b}_{NLLS})\,\text{Var}[\mathbf{b}_{NLLS}]\,\boldsymbol{G}(\mathbf{b}_{NLLS})'\}^{-1} R(\mathbf{b}_{NLLS})$$

$$\Rightarrow W \xrightarrow{d} \chi^2_J$$

NLLS - Application: A NIST Application (Greene)

Y	X
2.138	1.309
3.421	1.471
3.597	1.490
4.340	1.565
4.882	1.611
5.660	1.680

$$y = \beta_0 + \beta_1 x^{\beta 2} + \varepsilon.$$

$$x_i^0 = [1,\ x^{\beta 2},\ \beta_1 x^{\beta 2}\log x\,]$$

NLLS - Application: Iterations (Greene)

NLSQ; LHS=Y;
FCN=b0+B1*X^B2;
LABELS = b0, B1, B2;
MAXIT=500; TLF; TLB; OUTPUT=1; DFC ;
START=0,1,5 $

Begin NLSQ iterations. Linearized regression.
Iteration= 1; Sum of squares= 149.719219 ; Gradient= 149.718223
Iteration= 2; Sum of squares= 5.04072877 ; Gradient= 5.03960538
Iteration= 3; Sum of squares= .137768222E-01; Gradient= .125711747E-01
Iteration= 4; Sum of squares= .186786786E-01; Gradient= .174668584E-01
Iteration= 5; Sum of squares= .121182327E-02; Gradient= .301702148E-08
Iteration= 6; Sum of squares= .121182025E-02; Gradient= .134513256E-15
Iteration= 7; Sum of squares= .121182025E-02; Gradient= .644990175E-20
Convergence achieved

$$\text{Gradient} = [\mathbf{e}^0{}'\mathbf{X}^0]'[\mathbf{X}^0{}'\mathbf{X}^0]^{-1}\mathbf{X}^0{}'\mathbf{e}^0$$

NLLS - Application: Results (Greene)

```
-----------------------------------------------------------
User Defined Optimization.........................
Nonlinear     least squares regression ...........
LHS=Y         Mean                 =         4.00633
              Standard deviation   =         1.23398
              Number of observs.   =               6
Model size    Parameters           =               3
              Degrees of freedom   =               3
Residuals     Sum of squares       =          .00121
              Standard error of e  =          .02010
Fit           R-squared            =          .99984
--------+--------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er.  P[|Z|>z]
--------+--------------------------------------------------
      B0|    -.54559**        .22460       -2.429    .0151
      B1|    1.08072***       .13698        7.890    .0000
      B2|    3.37287***       .17847       18.899    .0000
--------+---------------------------------------------------
```

NLLS – Application: Solution (Greene)

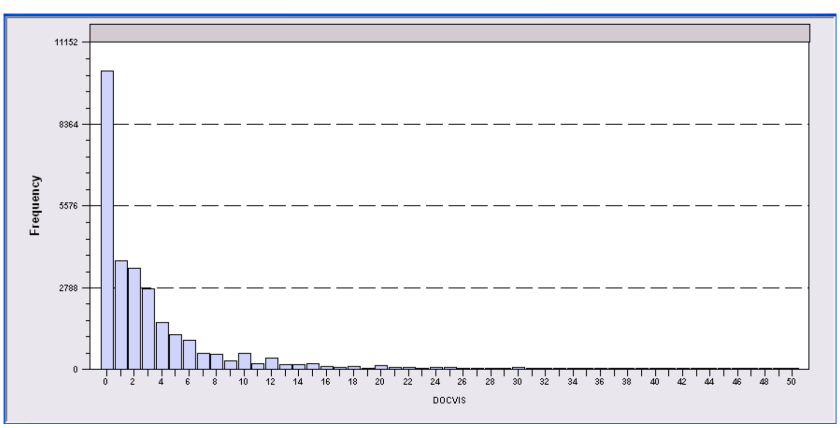
The pseudo regressors and residuals at the solution are:

X10	X20	X30	
1	x^{β_2}	$\beta_1 x^{\beta_2} \ln x$	e0
1	2.47983	0.721624	.0036
1	3.67566	1.5331	-.0058
1	3.83826	1.65415	-.0055
1	4.52972	2.19255	-.0097
1	4.99466	2.57397	.0298
1	5.75358	3.22585	-.0124

X0′e0 = .3375078D-13
.3167466D-12
.1283528D-10

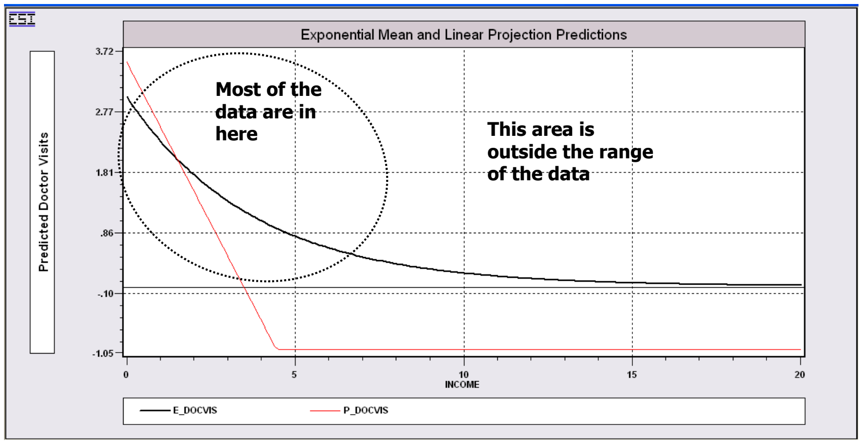
Application 2: Doctor Visits (Greene)

- German Individual Health Care data: N = 27,236
- Model for number of visits to the doctor
- Explanatory variables: Income, health, marital status, education, etc.



Application 2: Conditional Mean and Projection

• Plot: Number of visits to the doctor against household income:



Notice the problem with the linear approach. Negative predictions.

Application 2: NL Model Specification (Greene)

• Nonlinear Regression Model y = exp(**X**β) + ε

X = one, age, health_status, married, educ., household_income, nkids

- **nlsq;lhs=docvis;start=0,0,0,0,0,0,0;labels=k_b;fcn=exp(b1'x);maxit=25;out...**

Begin NLSQ iterations. Linearized regression.
Iteration= 1; Sum of squares= 1014865.00 ; Gradient= 257025.070
Iteration= 2; Sum of squares= .130154610E+11; Gradient= .130145942E+11
Iteration= 3; Sum of squares= .175441482E+10; Gradient= .175354986E+10
Iteration= 4; Sum of squares= 235369144. ; Gradient= 234509185.
Iteration= 5; Sum of squares= 31610466.6 ; Gradient= 30763872.3
Iteration= 6; Sum of squares= 4684627.59 ; Gradient= 3871393.70
Iteration= 7; Sum of squares= 1224759.31 ; Gradient= 467169.410
Iteration= 8; Sum of squares= 778596.192 ; Gradient= 33500.2809
Iteration= 9; Sum of squares= 746343.830 ; Gradient= 450.321350
Iteration= 10; Sum of squares= 745898.272 ; Gradient= .287180441
Iteration= 11; Sum of squares= 745897.985 ; Gradient= .929823308E-03
Iteration= **15**; Sum of squares= 745897.984 ; Gradient= .188041512E-10

Application 2: NL Regression Results (Greene)

```
| Nonlinear   least squares regression                   |
| LHS=DOCVIS   Mean                 =   3.183525     |
|              Standard deviation   =   5.689690     |
| WTS=none     Number of observs.   =      27326     |
| Model size   Parameters           =          7     |
|              Degrees of freedom   =      27319     |
| Residuals    Sum of squares       =   745898.0     |
|              Standard error of e  =   5.224584     |
| Fit          R-squared            =   .1567778     |
|              Adjusted R-squared   =   .1568087     |
| Info criter. LogAmemiya Prd. Crt. =   3.307006     |
|              Akaike Info. Criter. =   3.307263     |
| Not using OLS or no constant. Rsqd & F may be < 0. |
+----------------------------------------------------+
+---------+--------------+----------------+--------+---------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] |
+---------+--------------+----------------+--------+---------+
 B1          2.37667859      .06972582    34.086   .0000
 B2           .00809310      .00088490     9.146   .0000
 B3          -.21721398      .00313992   -69.178   .0000
 B4           .00371129      .02051147      .181   .8564
 B5          -.01096227      .00435601    -2.517   .0118
 B6          -.26584001      .05664473    -4.693   .0000
 B7          -.09152326      .02128053    -4.301   .0000
```

Partial Effects in the Nonlinear Model (Greene)

What are the slopes?

Conditional Mean Function = $E[y|\mathbf{x}] = \exp(\beta'\mathbf{x})$

Derivatives of the conditional mean are the partial effects

$$\frac{\partial E[y|\mathbf{x}]}{\partial \mathbf{x}} = \exp(\beta'\mathbf{x}) \times \beta$$

= a scaling of the coefficients that depends on the data

Usually computed using the sample means of the data.

Asymptotic Variance of the Slope Estimator (Greene)

$\hat{\delta}$ = estimated partial effects = $\dfrac{\partial \hat{E}[y|\mathbf{x}]}{\partial \mathbf{x}} \,|\, (\mathbf{x} = \bar{\mathbf{x}})$

To estimate Asy.Var[$\hat{\delta}$], we use the delta method:

$\hat{\delta} = \exp(\bar{\mathbf{x}}'\hat{\beta})\ \hat{\beta}$

$\hat{\mathbf{G}} = \dfrac{\partial \hat{\delta}}{\partial \hat{\beta}} = \exp(\bar{\mathbf{x}}'\hat{\beta})\ \mathrm{I} + \ \hat{\beta} \exp(\bar{\mathbf{x}}'\hat{\beta})\bar{\mathbf{x}}'$

Est.Asy.Var[$\hat{\delta}$]=$\hat{\mathbf{G}}$ Est.Asy.Var[$\hat{\beta}$] $\hat{\mathbf{G}}'$

Computing the Slopes (Greene)

```
calc;k=col(x)$
nlsq;lhs=docvis;start=0,0,0,0,0,0,0
    ;labels=k_b;fcn=exp(b1'x);
matr;xbar=mean(x)$
calc;mean=exp(xbar'b)$
matr;me=b*mean$
matr;g=mean*iden(k)+mean*b*xbar'$
matr;vme=g*varb*g'$
matr;stat(me,vme)$
```

Partial Effects at the Means of X (Greene)

```
----------------------------------------------------------
Number of observations in current sample =   27326
Number of parameters computed here       =       7
Number of degrees of freedom             =   27319
--------+--------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]
--------+--------------------------------------------------
Constant|    6.48148***       .20680       31.342   .0000
     AGE|     .02207***       .00239        9.216   .0000
    HSAT|    -.59241***       .00660      -89.740   .0000
 MARRIED|     .01005          .05593         .180   .8574
    EDUC|    -.02988**        .01186       -2.519   .0118
  HHNINC|    -.72495***       .15450       -4.692   .0000
  HHKIDS|    -.24958***       .05796       -4.306   .0000
--------+--------------------------------------------------
```

What About Just Using LS? (Greene)

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Least Squares Coefficient Estimates					
Constant	9.12437987	.25731934	35.459	.0000	
AGE	.02385640	.00327769	7.278	.0000	43.5256898
NEWHSAT	-.86828751	.01441043	-60.254	.0000	6.78566201
MARRIED	-.02458941	.08364976	-.294	.7688	.75861817
EDUC	-.04909154	.01455653	-3.372	.0007	11.3206310
HHNINC	-1.02174923	.19087197	-5.353	.0000	.35208362
HHKIDS	-.38033746	.07513138	-5.062	.0000	.40273000
Estimated Partial Effects					
ME_1	Constant term, marginal effect not computed				
ME_2	**.02207102**	**.00239484**	**9.216**	**.0000**	
ME_3	**-.59237330**	**.00660118**	**-89.737**	**.0000**	
ME_4	**.01012122**	**.05593616**	**.181**	**.8564**	
ME_5	**-.02989567**	**.01186495**	**-2.520**	**.0117**	
ME_6	**-.72498339**	**.15449817**	**-4.693**	**.0000**	
ME_7	**-.24959690**	**.05796000**	**-4.306**	**.0000**	

