Lecture 9
NLLS

1

M-Estimation

• An extremum estimator is one obtained as the optimizer of a criterion function, $q(\mathbf{z},\mathbf{b})$.

Examples:

OLS: $\mathbf{b}$ = arg max $(-\mathbf{e}'\mathbf{e}/T)$

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

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

• 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:

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

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

$$\sum_i q(\mathbf{z}_i,\mathbf{b})/T \xrightarrow{p} E[q(\mathbf{z},\boldsymbol{\beta})]$$

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

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

$$\text{Zero-Score:} \quad \sum_i \partial q(\mathbf{z}_i,\mathbf{b})/\partial \mathbf{b}' = \mathbf{0}$$

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

$$\mathbf{H} = \sum_i \partial^2 q(\mathbf{z}_i,\mathbf{b})/\partial \mathbf{b}\partial \mathbf{b}'$$

M-Estimation

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

$$\sum_i \mathbf{s}(\mathbf{z}_i,\mathbf{b}_M)/T = 0 \qquad (*)$$

• If, in addition, $E_X[\mathbf{s}(\mathbf{z},\mathbf{b})] = \partial/\partial \mathbf{b}'\, E_X[q(\mathbf{z},\mathbf{b})]$ -i.e., differentiation and integration are exchangeable-, then

$$E_X[\partial q(\mathbf{z},\boldsymbol{\beta})/\partial \boldsymbol{\beta}'] = \mathbf{0}.$$

• Under this assumptions the M-estimator is said to be of *ψ-type* ($\psi = \mathbf{s}(\mathbf{z},\mathbf{b})$=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*. ($\varrho = q(\mathbf{z},\boldsymbol{\beta})$).

M-Estimation: LS & ML

- Least Squares
  - DGP: $\mathbf{y} = f(\mathbf{x},\boldsymbol{\beta}) + \boldsymbol{\varepsilon}$, $\quad \mathbf{z} = [y,\mathbf{x}]$
  - $q(\mathbf{z};\boldsymbol{\beta}) = S(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = \sum_{i=1,\ldots,T} (y_i - f(\mathbf{x}_i;\boldsymbol{\beta}))^2$
  - Now, we move from population to sample moments
  - $q(\mathbf{z};\mathbf{b}) = S(\mathbf{b}) = \mathbf{e}'\mathbf{e} = \sum_{i=1,\ldots,T} (y_i - f(\mathbf{x}_i;\mathbf{b}))^2$
  - $\mathbf{b}_{NLLS} = \text{argmin } S(\mathbf{b})$

- Maximum Likelihood
  - Let $f(\mathbf{x}_i,\boldsymbol{\beta})$ be the pdf of the data.
  - $L(\mathbf{x},\boldsymbol{\beta}) = \Pi_{i=1,\ldots,T} f(\mathbf{x}_i;\boldsymbol{\beta})$
  - $\log L(\mathbf{x},\boldsymbol{\beta}) = \sum_{i=1,\ldots,T} \ln f(\mathbf{x}_i;\boldsymbol{\beta})$
  - Now, we move from population to sample moments
  - $q(\mathbf{z},\mathbf{b}) = -\log L(\mathbf{x},\mathbf{b})$
  - $\mathbf{b}_{MLE} = \text{argmin} - \log L(\mathbf{x};\mathbf{b})$

M-Estimation: Minimum L_p-estimators

• Minimum L_p-estimators

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

• Special cases:

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

$\mathbf{s}(\mathbf{z};\boldsymbol{\beta}) = \sum_i (x_i - \mathbf{b}_M) = 0 \quad \Rightarrow \mathbf{b}_M = \sum_i x_i/T$

– $p = 1$: 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 $\mathbf{s}(\mathbf{z},\mathbf{b}) = \partial q(\mathbf{z},\mathbf{b})/\partial \mathbf{b}'$ exists, we solve

$$\textstyle\sum_i \mathbf{s}(\mathbf{z}_i,\mathbf{b}_M)/T = 0 \qquad (\mathbf{s}(\mathbf{z}_i,\mathbf{b}) \text{ is a } k\text{x1 vector}).$$

– $E[\mathbf{s}(\mathbf{z},\mathbf{b_0})] = E[\partial q(\mathbf{z},\mathbf{b})/\partial \mathbf{b}'] = \mathbf{0}$

– Using the LLN: $\sum_i \mathbf{s}(\mathbf{z}_i,\mathbf{b})/T \xrightarrow{p} E[\mathbf{s}(\mathbf{z},\mathbf{b_0})] = \mathbf{0}$

– $\boldsymbol{V} = \mathrm{Var}[\mathbf{s}(\mathbf{z},\mathbf{b_0})] = E[\mathbf{s}(\mathbf{z},\mathbf{b})\mathbf{s}(\mathbf{z},\mathbf{b})']$ ($\boldsymbol{V}$ is a kxk matrix).

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

– Using the LLN: $\sum_i [\mathbf{s}(\mathbf{z}_i,\mathbf{b})\mathbf{s}(\mathbf{z}_i,\mathbf{b})']/T \xrightarrow{p} \mathrm{Var}[\mathbf{s}(\mathbf{z},\mathbf{b_0})]$

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

<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: $\sum_i [\partial \mathbf{s}(\mathbf{z}_i,\mathbf{b})/\partial \mathbf{b}]/T \xrightarrow{p} \boldsymbol{H}(\mathbf{z},\mathbf{b_0})$

- 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.

– Recall the Mean Value Theorem: $f(x) = f(a) + f'(b)\,(x-a) \qquad a<b<x$

– Apply MVT to the score:

$$\textstyle\sum_i \mathbf{s}(\mathbf{z}_i,\mathbf{b_M}) = \sum_i \mathbf{s}(\mathbf{z}_i,\mathbf{b_0}) + \sum_i \mathbf{H}(\mathbf{z}_i,\mathbf{b^*})\,(\mathbf{b_M}-\mathbf{b_0}) \qquad \mathbf{b_0}<\mathbf{b^*}<\mathbf{b_M}$$

$$0 = \textstyle\sum_i \mathbf{s}(\mathbf{z}_i,\mathbf{b_0}) + \sum_i \mathbf{H}(\mathbf{z}_i,\mathbf{b^*})\,(\mathbf{b_M}-\mathbf{b_0})$$

$$\Rightarrow (\mathbf{b_M}-\mathbf{b_0}) = [\textstyle\sum_i \mathbf{H}(\mathbf{z}_i,\mathbf{b^*})]^{-1} \sum_i \mathbf{s}(\mathbf{z}_i,\mathbf{b_0})$$

$$\Rightarrow \sqrt{T}\,(\mathbf{b_M}-\mathbf{b_0}) = [\textstyle\sum_i \mathbf{H}(\mathbf{z}_i,\mathbf{b^*})/T]^{-1} \sum_i \mathbf{s}(\mathbf{z}_i,\mathbf{b_0})/\sqrt{T}$$

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, $\mathbf{b_M} \xrightarrow{p} \mathbf{b_0}$

Remark: 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 $E_X[q(\mathbf{z}, \mathbf{b_0})] = \min_{\mathbf{b}\in\mathbf{B}} E_X[q(\mathbf{z},\boldsymbol{\beta})]$.

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) $\mathbf{b_M} \xrightarrow{p} \mathbf{b_0}$ for some $\mathbf{b_0} \in \mathbf{B}$

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

(3) $\sum_i [\partial\mathbf{s}(\mathbf{z}_i,\mathbf{b})/\partial\mathbf{b}]/T|_{\mathbf{b}=\mathbf{b}^*} \xrightarrow{p} \boldsymbol{H}(\mathbf{z},\mathbf{b_0})$ for $\mathbf{b}^* \xrightarrow{p} \mathbf{b_0}$

(4) $\sum_i \mathbf{s}(\mathbf{z}_i,\mathbf{b})/\sqrt{T} \xrightarrow{d} N(\mathbf{0},\boldsymbol{V_0})$ $\boldsymbol{V_0} = \text{Var}[\mathbf{s}(\mathbf{z},\mathbf{b_0})] < \infty$

Then, $\sqrt{T}\,(\mathbf{b_M} - \mathbf{b_0}) = \{\sum_i \mathbf{H}(\mathbf{z}_i,\mathbf{b}^*)/T\}^{-1}[-\sum_i \mathbf{s}(\mathbf{z}_i,\mathbf{b_0})/\sqrt{T}]$

$\Rightarrow \sqrt{T}\,(\mathbf{b_M} - \mathbf{b_0}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{H_0}^{-1} \boldsymbol{V_0} \boldsymbol{H_0}^{-1})$

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

Asymptotic Normality

- Summary
  - $\mathbf{b_M} \xrightarrow{p} \mathbf{b_0}$
  - $\mathbf{b_M} \xrightarrow{a} N(\mathbf{b_0}, \text{Var}[\mathbf{b_0}])$
  - $\text{Var}[\mathbf{b_M}] = (1/T)\, \boldsymbol{H_0}^{-1} \boldsymbol{V_0} \boldsymbol{H_0}^{-1}$
  - If the model is correctly specified: $-\mathbf{H} = \mathbf{V}$.
    Then, $\text{Var}[\mathbf{b}] = \boldsymbol{V_0}$

  – $\mathbf{H}$ and $\mathbf{V}$ are evaluated at $\mathbf{b_0}$:
    - $\mathbf{H} = \sum_i [\partial^2 q(\mathbf{z}_i, \mathbf{b})/\partial \mathbf{b} \partial \mathbf{b}']$
    - $\mathbf{V} = \sum_i [\partial q(\mathbf{z}_i, \mathbf{b})/\partial \mathbf{b}][\partial q(\mathbf{z}_i, \mathbf{b})/\partial \mathbf{b}']$

M-Estimation: Example

- DGP: $\mathbf{y} = f(\mathbf{x}_i;\boldsymbol{\beta}) + \boldsymbol{\varepsilon} = \exp(\mathbf{x}\boldsymbol{\beta}) + \boldsymbol{\varepsilon}$,
  - Objective function:
    $q(\mathbf{X},\boldsymbol{\beta}) = ½\, \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon} = ½\, [\mathbf{y} - \exp(\mathbf{X}\boldsymbol{\beta})]' [\mathbf{y} - \exp(\mathbf{X}\boldsymbol{\beta})]$

  - Score: $\mathbf{s}(\mathbf{z},\boldsymbol{\beta}) = \partial q(\mathbf{z},\boldsymbol{\beta})/\partial \boldsymbol{\beta} = \partial f(\mathbf{x}_i;\boldsymbol{\beta})/\partial \boldsymbol{\beta}'\, \boldsymbol{\varepsilon}$
    $= - [\exp(\mathbf{X}\boldsymbol{\beta})\mathbf{X}]' [\mathbf{y} - \exp(\mathbf{X}\boldsymbol{\beta})]$
    $= - [\exp(\mathbf{X}\boldsymbol{\beta})\mathbf{X}]' \boldsymbol{\varepsilon} = - \mathbf{X}' \exp(\mathbf{X}\boldsymbol{\beta})' \mathbf{y} + \mathbf{X}' \exp(2\sum_i \mathbf{x_i}'\boldsymbol{\beta})$

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

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

  - $\text{Var}[\mathbf{b_M}] = (1/T)\, \boldsymbol{H_0}^{-1} \boldsymbol{V_0} \boldsymbol{H_0}^{-1}$

M-Estimation: Example

• $\text{Var}[\mathbf{b_M}] = (1/T)\ \boldsymbol{H_0}^{-1}\, \boldsymbol{V_0}\, \boldsymbol{H_0}^{-1}$

• We approximate

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

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

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: $\mathbf{y_2} = g(\mathbf{w};\boldsymbol{\gamma}) + \boldsymbol{\nu}$ => we estimate $\boldsymbol{\gamma}$, say $\mathbf{c}$

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

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

• <u>Examples</u>:

(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 **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 $\sqrt{T}\,(\mathbf{b_M} - \mathbf{b_0}) = \boldsymbol{H}_0^{-1}[-\sum_i \mathbf{s}(\mathbf{z}_i,\mathbf{b_0},\mathbf{c})/\sqrt{T}] + o(1)$ (*)

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 $\boldsymbol{F}_0 = \partial\mathbf{s}(\mathbf{z},\mathbf{b_0},\mathbf{c})/\partial\boldsymbol{\gamma}$

<u>Note</u>: If $\mathbf{c}=\mathbf{c_0}$ or $\boldsymbol{F}_0 = 0$, then (**) holds.

Two-Step M-Estimation

• We can also write $\sqrt{T}\,(\mathbf{c} - \mathbf{c_0}) = \boldsymbol{H}_{\mathbf{c0}}^{-1}[-\sum_i \mathbf{s}(\mathbf{w}_i,\mathbf{c})/\sqrt{T}] + o(1)$
$= \sum_i \mathbf{h}(\mathbf{w}_i,\mathbf{c})/\sqrt{T} + o(1)$

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

$$\sqrt{T}\,(\mathbf{b_M} - \mathbf{b_0}) = \boldsymbol{H}_\mathbf{0}^{-1}[-\sum_i \mathbf{r}(\mathbf{z}_i,\mathbf{b_0},\mathbf{c_0})/\sqrt{T}] + o(1), \qquad (****)$$

where $\mathbf{r}(\mathbf{z}_i,\mathbf{b_0},\mathbf{c_0}) = \mathbf{s}(\mathbf{z}_i,\mathbf{b_0},\mathbf{c_0}) + \boldsymbol{F}_\mathbf{0}\ \mathbf{h}(\mathbf{w}_i,\mathbf{c_0})$

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 1^{st}-stage adjustment.

• $\mathrm{Var}[\mathbf{b_M}] = (1/T)\ \boldsymbol{H}_\mathbf{0}^{-1}\ \mathbf{Var}[\mathbf{r}(\mathbf{z}_i,\mathbf{b_0},\mathbf{c_0})]\ \boldsymbol{H}_\mathbf{0}^{-1}$

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: $y_i = f(\mathbf{x}_i, \boldsymbol{\beta}) + \varepsilon_i$

• 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:

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

Example: A non intrinsic linear model:

$$y_i = \beta_0 + \beta_1 x_i^{\beta 2} + \varepsilon_i.$$

Nonlinear Least Squares

• Least squares: $\text{Min}_{\boldsymbol{\beta}} \; S(\boldsymbol{\beta}) = \{½ \Sigma_i [y_i - f(\mathbf{x}_i,\boldsymbol{\beta})]^2\} = ½ \Sigma_i \varepsilon_i^2$

F.o.c.:

$$\partial ½\Sigma_i[y_i - f(\mathbf{x}_i,\boldsymbol{\beta})]^2]/\partial\boldsymbol{\beta}$$

$$= ½\Sigma_i(-2)[y_i - f(\mathbf{x}_i,\boldsymbol{\beta})] \, \partial f(\mathbf{x}_i,\boldsymbol{\beta})/\partial\boldsymbol{\beta} = -\Sigma_i \, e_i \, \mathbf{x_i^0}$$

$$\Rightarrow \; -\Sigma_i \, e_i \, \mathbf{x_i^0} = \mathbf{0} \qquad \text{we solve for } \mathbf{b}_{NLLS}$$

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

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

• 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: $\text{Min}_{\beta}\ S(\boldsymbol{\beta}) = \{½\ \Sigma_i\ [y_i\ - f(\mathbf{x_i},\boldsymbol{\beta})]^2\ \} = ½\ \Sigma_i\ \varepsilon_i^2$

$y_i\ = f(\mathbf{x_i},\boldsymbol{\beta}) + \varepsilon_i = \beta_0 + \beta_1\, x_i^{\beta 2} + \varepsilon_i.$

f.o.c.:

$\partial[\ ½\Sigma_i\ e_i^2]/\partial\beta_0\ = \Sigma_i\ (-1)\ (y_i - \beta_0 - \beta_1\, x_i^{\beta 2})\ 1 = 0$

$\partial[\ ½\Sigma_i\ e_i^2]/\partial\beta_1\ = \Sigma_i\ (-1)\ (y_i - \beta_0 - \beta_1\, x_i^{\beta 2})\ x_i^{\beta 2} = 0$

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

• 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. We would get the normal equations from the f.o.c.

Nonlinear Least Squares: Example

Example: $\text{Min}_{\beta}\ S(\boldsymbol{\beta}) = \{½\ \Sigma_i\ [y_i\ - (\beta_0 + \beta_1\, x_i^{\beta 2})]^2\ \}$

• 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 $\boldsymbol{\beta}^0$ => get RSS($\boldsymbol{\beta}^0$)
3. Make small changes to $\boldsymbol{\beta}^0$, => get $\boldsymbol{\beta}^1$.
4. Calculate RSS for $\boldsymbol{\beta}^1$ => get RSS($\boldsymbol{\beta}^1$)
5. If RSS($\boldsymbol{\beta}^1$) < RSS($\boldsymbol{\beta}^0$) => $\boldsymbol{\beta}^1$ becomes your new starting point.
6. Repeat steps 3-5 until you RSS($\boldsymbol{\beta}^j$) cannot be lowered. => get $\boldsymbol{\beta}^j$.

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

NLLS: Linearization

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

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

$$f(\mathbf{x}_i, \boldsymbol{\beta}) \approx f(\mathbf{x}_i, \boldsymbol{\beta}^0) + \Sigma_k[\partial f(\mathbf{x}_i, \boldsymbol{\beta}^0)/\partial \beta_k^0](\beta_k - \beta_k^0)$$
$$= f(\mathbf{x}_i, \boldsymbol{\beta}^0) + \Sigma_k \mathbf{x}_i^0 (\beta_k - \beta_k^0)$$
$$= [f(\mathbf{x}_i, \boldsymbol{\beta}^0) - \Sigma_k \mathbf{x}_i^0 \beta_k^0] + \Sigma_k \mathbf{x}_i^0 \beta_k$$
$$= f^0 + \Sigma_k \mathbf{x}_i^0 \beta_k = f^0 + \mathbf{x}_i^{0\prime} \boldsymbol{\beta}$$

where

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

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

$$y_i = f_i^0 + \mathbf{x}_i^{0\prime} \boldsymbol{\beta} + \boldsymbol{\varepsilon}^0_i \qquad (\boldsymbol{\varepsilon}^0_i = \boldsymbol{\varepsilon}_i + \text{linearization error i})$$
$$\Rightarrow y^0_i = y_i - f_i^0 = \mathbf{x}_i^{0\prime} \boldsymbol{\beta} + \boldsymbol{\varepsilon}^0_i$$

NLLS: Linearization

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

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

• Now, we can do OLS:

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

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

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

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

NLLS: Linearization

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

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

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

• 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: $y_i = f(\mathbf{x}_i, \boldsymbol{\beta}^0) + \varepsilon_i = \beta_0 + \beta_1 x_i^{\beta 2} + \varepsilon_i$

• Linearize the model to get:

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

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

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

$\partial f(\mathbf{x}_i, \boldsymbol{\beta})/\partial \beta_1 = x_i^{\beta 2}$

$\partial f(\mathbf{x}_i, \boldsymbol{\beta})/\partial \beta_2 = \beta_1 x_i^{\beta 2} \ln(\mathbf{x}_i)$

$$f_i^0 = \beta^0_0 + \beta^0_1 x_i^{\beta 0,2} - \{\beta^0_0 + \beta^0_1 x_i^{\beta 0,2} + \beta^0_2 \beta^0_1 x_i^{\beta 0,2} \ln(x_i)\}$$

$$y^0_i = \beta_0 + \beta_1 x_i^{\beta 0,2} + \beta_2 \beta^0_1 \mathbf{x}_i^{\beta 0,2} \ln(x_i) + \varepsilon^0_i$$

To get $\mathbf{b}_{NLLS}$, regress $\mathbf{y}^0$ on a constant, $\mathbf{x}^{\beta 0,2}$, and $\beta^0_1 \mathbf{x}^{\beta 0,2} \ln(\mathbf{x})$.

Gauss-Newton Algorithm

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

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

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

$$\boldsymbol{\beta}_{k+1} = \boldsymbol{\beta}_k + (\mathbf{J}^T\,\mathbf{J})^{-1}\,\mathbf{J}^T\,\boldsymbol{\varepsilon} \qquad \text{-- } \mathbf{J}\text{: Jacobian} = \delta f(xi;\boldsymbol{\beta})/\delta\boldsymbol{\beta}.$$

- Given a $\mathbf{b}_{NLLS}$ at step m, $\mathbf{b}(j)$, 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: $\partial f(x_i,\mathbf{b}(j))/\partial \mathbf{b}(j)'$

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

- 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.

$$\mathbf{y} = f(\mathbf{x}_i,\boldsymbol{\beta}) + \boldsymbol{\varepsilon} = \Sigma_k\,\mathbf{x}_k^{(\lambda)}\,\beta_k + \boldsymbol{\varepsilon}$$

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

- 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:

$$\mathbf{y}^{(\lambda 1)} = \Sigma_k\,\mathbf{x}_k^{(\lambda 2)}\,\beta_k + \boldsymbol{\varepsilon}$$

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 rank$[\partial R(\boldsymbol{\beta})/\partial\boldsymbol{\beta}=\mathbf{G}(\boldsymbol{\beta})]=J$.

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

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

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

Solution: 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,

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

$$\Rightarrow W \xrightarrow{d} \chi_J^2$$

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;<u>DFC</u> ;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)

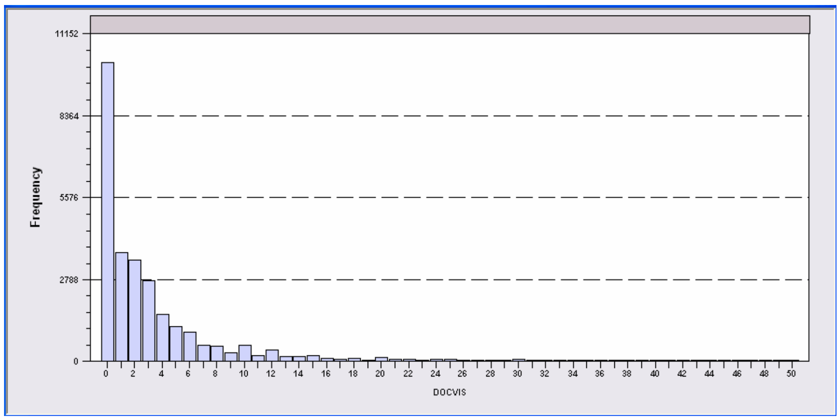
The pseudo regressors and residuals at the solution are:

X10	X20	X30	
1	$x^{\beta 2}$	$\beta_1 x^{\beta 2}$ lnx	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

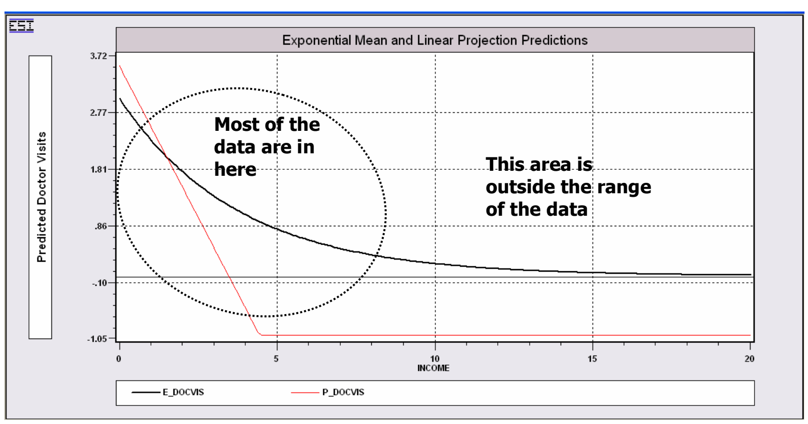
X0′e0 = .3375078D-13
.3167466D-12
.1283528D-10

Application 2: Doctor Visits (Greene)

- German Individual Health Care data: N=27,236
- Model for number of visits to the doctor



Application 2: Conditional Mean and Projection



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

Derivatives of the conditional mean are the partial effects

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

= 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)

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

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

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

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

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

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**	