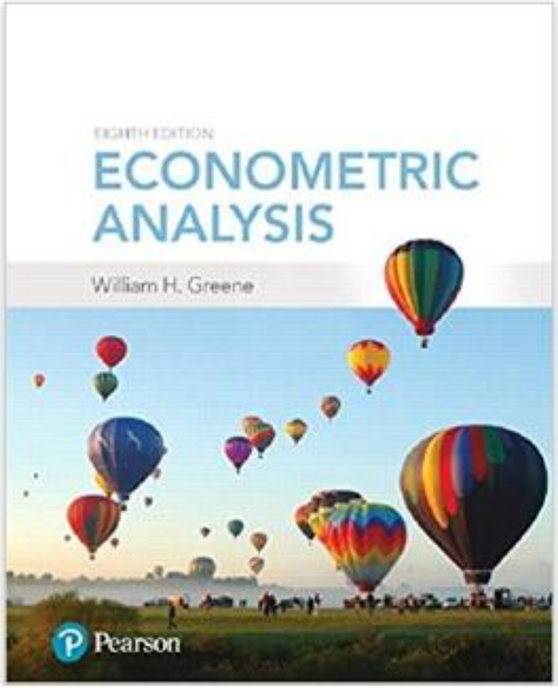
Econometrics I

Professor William Greene
Stern School of Business
Department of Economics



Econometrics I

Part 17 – Nonlinear Regression

Nonlinear Regression

- Nonlinearity and Nonlinear Models
- Estimation Criterion
- Iterative Algorithm

Nonlinear Regression

What makes a regression model "nonlinear?"

Nonlinear functional form?

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

Not necessarily: $y_i = \exp(\alpha) + \beta_2 * \log(x_i) + \varepsilon_i$

$\beta_1 = \exp(\alpha)$

$y_i = \exp(\alpha) x_i^{\beta_2} \exp(\varepsilon_i)$ is "loglinear"

Models can be nonlinear in the functional form of the relationship between y and x, and not be nonlinear for purposes here.

We will redefine "nonlinear" shortly, as we proceed.

Least Squares

Least squares: Minimize wrt $\boldsymbol{\beta}$

$$S(\boldsymbol{\beta}) = \tfrac{1}{2} \Sigma_i \{y_i - E[y_i|\mathbf{x}_i,\boldsymbol{\beta}]\}^2$$

$$= \tfrac{1}{2} \Sigma_i [y_i - f(\mathbf{x}_i,\boldsymbol{\beta})]^2$$

$$= \tfrac{1}{2} \Sigma_i e_i^2$$

First order conditions: $\partial S(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = \mathbf{0}$

$$\partial\{\tfrac{1}{2}\Sigma_i[y_i - f(\mathbf{x}_i,\boldsymbol{\beta})]^2\}/\partial\boldsymbol{\beta}$$

$$= \tfrac{1}{2}\Sigma_i(-2)[y_i - f(\mathbf{x}_i,\boldsymbol{\beta})]\ \partial f(\mathbf{x}_i,\boldsymbol{\beta})/\partial\boldsymbol{\beta}$$

$$= -\Sigma_i e_i \mathbf{x}_i^0 = \mathbf{0} \quad \text{(familiar?)}$$

There is no explicit solution, **b** = f(**data**) like LS.

(Nonlinearity of the FOC defines nonlinear model)

Example

How to solve this kind of set of equations: Example,

$$y_i = \beta 0 + \beta 1 x_i^{\beta 2} + \varepsilon_i.$$

$$\partial[\tfrac{1}{2}\Sigma_i e_i^2]/\partial\beta 0 = \Sigma_i (-1)(y_i - \beta 0 - \beta 1 x_i^{\beta 2})\ 1 = 0$$
$$\partial[\tfrac{1}{2}\Sigma_i e_i^2]/\partial\beta 1 = \Sigma_i (-1)(y_i - \beta 0 - \beta 1 x_i^{\beta 2})\ x_i^{\beta 2} = 0$$
$$\partial[\tfrac{1}{2}\Sigma_i e_i^2]/\partial\beta 2 = \Sigma_i (-1)(y_i - \beta 0 - \beta 1 x_i^{\beta 2})\ \beta 1 x_i^{\beta 2} \ln x_i = 0$$

Nonlinear equations require a nonlinear solution.

This defines a nonlinear regression model.

I.e., when the first order conditions are not linear in $\boldsymbol{\beta}$.

(!!!) Check your understanding. What does this produce if $f(\mathbf{x}_i,\boldsymbol{\beta}) = \mathbf{x}_i'\boldsymbol{\beta}$? (I.e., a linear model)

The Linearized Regression Model

Linear Taylor series: $y = f(\mathbf{x}_i,\boldsymbol{\beta}) + \varepsilon$.

Expand the regression around some point, $\boldsymbol{\beta}^*$.

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

$$= f(\mathbf{x}_i,\boldsymbol{\beta}^*) + \Sigma_k x_{ik}^0(\beta_k - \beta_k^*)$$

$$= [f(\mathbf{x}_i,\boldsymbol{\beta}^*) - \Sigma_k x_{ik}^0\beta_k^*] + \Sigma_k x_i^0\beta_k$$

$$= f^0 + \Sigma_k x_{ik}^0\beta_k \text{ which looks linear.}$$

x_{ik}^0 = the derivative wrt β_k evaluated at $\boldsymbol{\beta}^*$

The '**pseudo-regressors**' are the derivative functions in the linearized model.

Estimating Asy.Var[b]

Computing the asymptotic covariance matrix for the nonlinear least squares estimator using the pseudo regressors and the sum of squares.

$$\text{Est.Asy.Var}[\mathbf{b}] = \hat{\sigma}^2\left[\sum_{i=1}^{n}\mathbf{x}_i^0(\mathbf{b})\mathbf{x}_i^0(\mathbf{b})'\right]^{-1} = \hat{\sigma}^2\left[\mathbf{X}(\mathbf{b})^0{}'\mathbf{X}(\mathbf{b})^0\right]^{-1}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f(\mathbf{x}_i, \mathbf{b})\right)^2 \text{ not } \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \mathbf{x}_i^0{}'\mathbf{b}\right)^2$$

(I.e., deviations from estimated regression, not estimated linearized regression.) Often "degrees of freedom" corrected

$$\hat{\sigma}^2 = \frac{1}{n-K}\sum_{i=1}^{n}\left(y_i - f(\mathbf{x}_i, \mathbf{b})\right)^2 \quad K = \#\text{ of parameters in } \beta.$$

Gauss-Marquardt Algorithm

Given a coefficient vector **b**(m), at step m, find the vector for step m+1 by

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

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

$\mathbf{e}^0$ = vector of residuals, $\mathbf{y} - \mathbf{f}[\mathbf{x},\mathbf{b}(m)]$

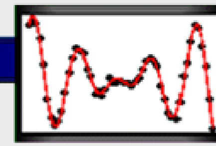
"Update" vector is the slopes in the regression of the residuals on the pseudo-regressors. Update is zero when they are orthogonal. (Just like LS)

StRD Nonlinear Least Squares Regression Datasets - Windows Internet Explorer

http://www.itl.nist.gov/div898/strd/nls/nls_main.shtml

NIST StRD

Nonlinear Regression



[Background Information](#)

Dataset Name	Level of Difficulty	Model Classification	Number of Parameters	Number of Observations	Source
Misra1a	Lower	Exponential	2	14	Observed
Chwirut2	Lower	Exponential	3	54	Observed
Chwirut1	Lower	Exponential	3	214	Observed
Lanczos3	Lower	Exponential	6	24	Generated
Gauss1	Lower	Exponential	8	250	Generated
Gauss2	Lower	Exponential	8	250	Generated
DanWood	Lower	Miscellaneous	2	6	Observed
Misra1b	Lower	Miscellaneous	2	14	Observed
Kirby2	Average	Rational	5	151	Observed

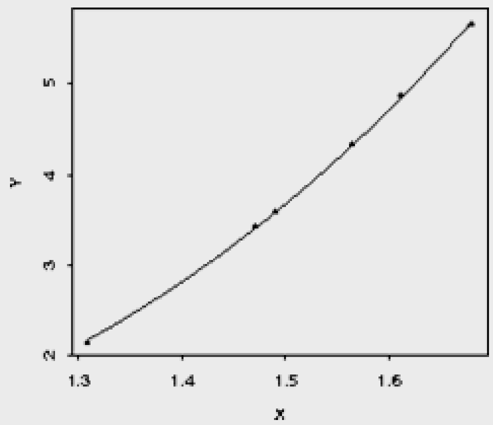
[Main NLS Page](#) [Previous Dataset](#) [Next Dataset](#)

Dataset Information

Dataset Name: DanWood

Procedure: Nonlinear Least Squares Regression
[Certification Method & Definitions](#)

Data: 1 Response Variable (y)
1 Predictor Variable (x)
6 Observations
Lower Level of Difficulty
Observed Data
[Data file (ASCII Format)](#)
[Additional Information](#)



Model: Miscellaneous Class
2 Parameters (β_1, β_2)
[Starting Values](#)

$$y = f(x;\beta) + \epsilon$$

$$= \beta_1 x^{\beta_2} + \epsilon$$

Results: [Certified Values](#)
[Graphics](#)

Reference: Daniel, C. and F. S. Wood (1980).
Fitting Equations to Data, Second Edition.
New York, NY: John Wiley and Sons, pp. 428-431.

The NIST Dan Wood Application

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

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

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

Dan Wood Solution

```
Description:   These data and model are described in Daniel and Wood
               (1980), and originally published in E.S.Keeping,
               "Introduction to Statistical Inference," Van Nostrand
               Company, Princeton, NJ, 1962, p. 354.  The response
               variable is energy radieted from a carbon filament
               lamp per cm**2 per second, and the predictor variable
               is the absolute temperature of the filament in 1000

               degrees Kelvin.
Reference:     Daniel, C. and F. S. Wood (1980).
               Fitting Equations to Data, Second Edition.
               New York, NY:  John Wiley and Sons, pp. 428-431.
Data:          1 Response Variable  (y = energy)
               1 Predictor Variable (x = temperature)
               6 Observations
               Lower Level of Difficulty
Model:         Miscellaneous Class
               y  = b1*x**b2  +  e
          Starting values                         Certified Values

        Start 1     Start 2               Parameter     Standard Deviation
  b1 =    1           0.7           7.6886226176E-01  1.8281973860E-02
  b2 =    5           4             3.8604055871E+00  5.1726610913E-02
Residual Sum of Squares:                    4.3173084083E-03
Residual Standard Deviation:                3.2853114039E-02
Degrees of Freedom:                                4
```

Iterations

NLSQ;LHS=Y ;FCN=B1*X^B2 ;LABELS=B1,B2
;MAXIT=500;TLF;TLB;OUTPUT=1;<u>DFC </u>;START=1,5 $

Begin NLSQ iterations. Linearized regression.
Iteration= 1; Sum of squares= 149.719219 ; Gradient = 149.717729
Iteration= 2; Sum of squares= 4.61087813 ; Gradient = 4.60758767
Iteration= 3; Sum of squares= .316841645E-01; Gradient = .273915086E-01
Iteration= 4; Sum of squares= .432059390E-02; Gradient = .328593114E-05
Iteration= 5; Sum of squares= .431730841E-02; Gradient = .383744563E-11
Iteration= 6; Sum of squares= .431730841E-02; Gradient = .225037906E-15
Iteration= 7; Sum of squares= .431730841E-02; Gradient = .140146409E-19
Iteration= 8; Sum of squares= .431730841E-02; Gradient = .162996996E-23
Convergence achieved

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

Results

```
-----------------------------------------------------------------------------
User Defined Optimization.........................
Nonlinear    least squares regression ............
LHS=Y        Mean                    =       4.00633
             Standard deviation      =       1.23398
             Number of observs.      =             6
Model size   Parameters              =             2
             Degrees of freedom      =             4
Residuals    Sum of squares          =    .431731E-02
             Standard error of e     =        .03285
Fit          R-squared               =        .99943
             Adjusted R-squared      =        .99929
Model test   F[  1,      4] (prob) =  7050.0(.0000)
Not using OLS or no constant. Rsqrd & F may be < 0
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
UserFunc|  Coefficient     Error        z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
      B1|     .76886***    .01828    42.06  .0000      .73303    .80469
      B2|    3.86041***    .05173    74.63  .0000     3.75902   3.96179
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

NLS Solution

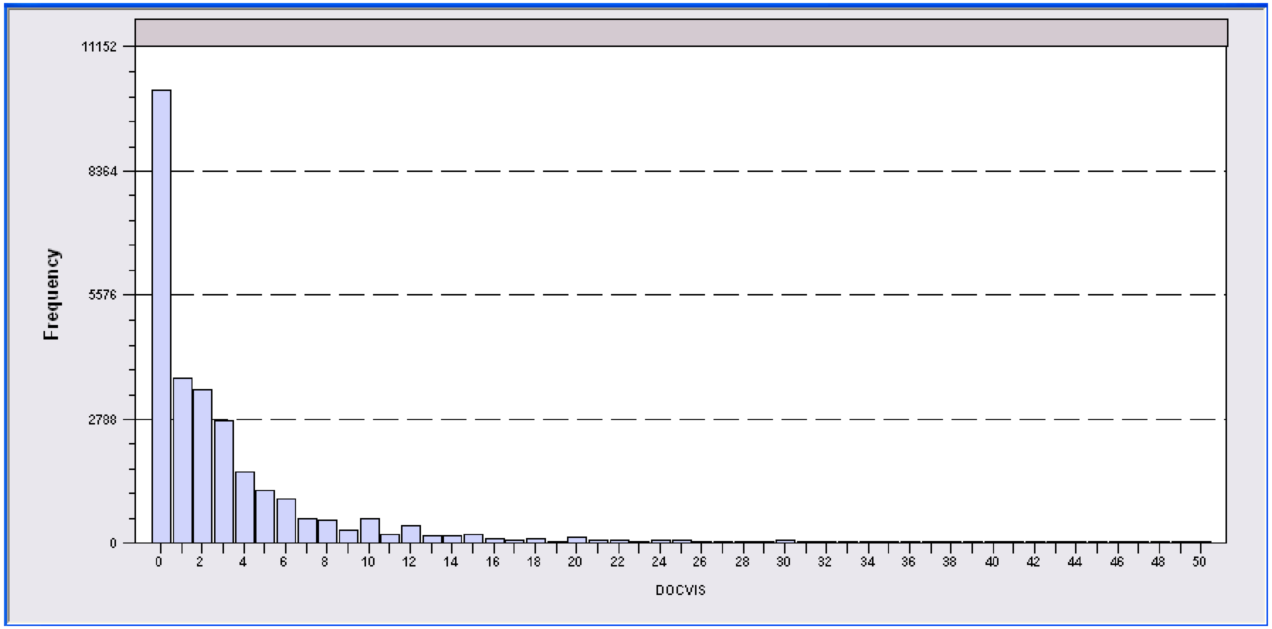
The pseudo regressors and residuals at the solution are:

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

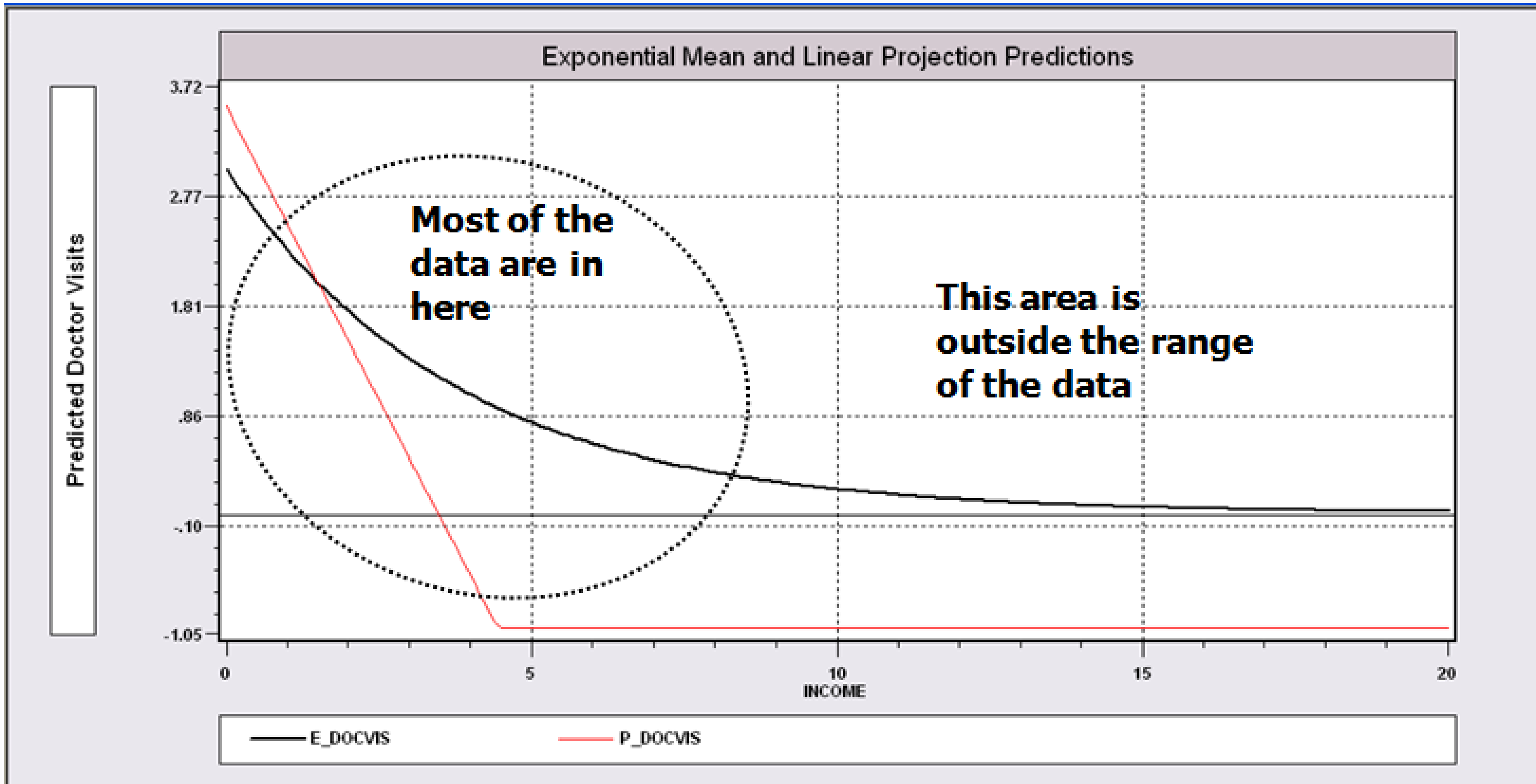
X0′e0 = .3375078D-13
.3167466D-12
.1283528D-10

Application: Doctor Visits

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



Conditional Mean and Projection



Notice the problem with the linear approach. Negative predictions.

Nonlinear Model Specification

Nonlinear Regression Model $y=\exp(\boldsymbol{\beta}'\mathbf{x}) + \varepsilon$

X =one,age,health_status, married,

education, household_income, kids

NLS Iterations

```
--> 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= 12; Sum of squares=  745897.984       ; Gradient=  .839914514E-05
Iteration= 13; Sum of squares=  745897.984       ; Gradient=  .991471058E-07
Iteration= 14; Sum of squares=  745897.984       ; Gradient=  .132954206E-08
Iteration= 15; Sum of squares=  745897.984       ; Gradient=  .188041512E-10
```

(The blue iterations take place within the ‘hidden digits.’)

Nonlinear Regression Results

```
-----------------------------------------------------------------------------
User Defined Optimization.........................
Nonlinear     least squares regression ............
LHS=DOCVIS    Mean                 =         3.18352
              Standard deviation   =         5.68969
              Number of observs.   =           27326
Model size    Parameters           =               7
              Degrees of freedom   =           27319
Residuals     Sum of squares       =         745892.
              Standard error of e  =         5.22456
Fit           R-squared            =          .15678
              Adjusted R-squared   =          .15682
Model test    F[  6, 27319] (prob) =     846.6(.0000)
Diagnostic    Log likelihood       =    -83953.92816
              Restricted(b=0)      =    -86283.92356
              Chi-sq [   6]  (prob) =4660.0(  .0000)
Info criter. Akaike Info. Criter. =         3.30725
Not using OLS or no constant. Rsqrd & F may be < 0
```

UserFunc	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
B1	2.37668***	.06972	34.09	.0000	2.24003	2.51334
B2	.00809***	.00088	9.15	.0000	.00636	.00983
B3	-.21723***	.00314	-69.18	.0000	-.22338	-.21107
B4	.00369	.02051	.18	.8574	-.03652	.04389
B5	-.01096**	.00436	-2.52	.0119	-.01949	-.00242
B6	-.26583***	.05664	-4.69	.0000	-.37685	-.15481
B7	-.09152***	.02128	-4.30	.0000	-.13323	-.04981

Partial Effects in the Nonlinear Model

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

(1) Computed using the sample means of the data.

PEA = Partial Effects at Averages

(2) Computed at each observation then averaged

APE = Average Partial Effects

Delta Method for Asymptotic Variance of the Slope Estimator

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

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

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

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

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

To compute Average Partial Effects :

(1) Use $\bar{\hat{\delta}} = (1/n)\sum_{i=1} \hat{\delta}(\mathbf{x}_i)$

(2) Use $\bar{\hat{\mathbf{G}}} = (1/n)\sum_{i=1} \hat{\mathbf{G}}(\mathbf{x}_i)$

Computing the Slopes

```
Namelist;x=one,age,hsat,married,educ,hhninc,hhkids$
Calc      ; k=col(x)$
Nlsq      ; lhs=docvis;start=0,0,0,0,0,0,0
          ; labels=k_b;fcn=exp(b1'x)$
Partials  ; function=exp(b1'x)
          ; parameters=b
          ; covariance = varb
          ; labels=k_b
          ; effects : x ; summary [; means] $
```

Partial Effects: PEA vs. APE

```
-----------------------------------------------------------------------
Partial Effects for User Specified Function
Partial Effects Computed at data Means
* ==> Partial Effect for a Binary Variable
-----------------------------------------------------------------------
                     Partial     Standard
(Delta method)       Effect       Error      |t|   95% Confidence Interval
-----------------------------------------------------------------------
      AGE            .02207      .00239     9.22      .01738      .02676
      HSAT         -.59241      .00660    89.74    -.60535    -.57947
      MARRIED        .01005      .05593      .18    -.09958      .11968
      EDUC         -.02988      .01186     2.52    -.05314    -.00663
      HHNINC       -.72494      .15450     4.69   -1.02775    -.42214
      HHKIDS       -.24958      .05796     4.31    -.36318    -.13599
-----------------------------------------------------------------------
|
-----------------------------------------------------------------------
Partial Effects for User Specified Function
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
-----------------------------------------------------------------------
                     Partial     Standard
(Delta method)       Effect       Error      |t|   95% Confidence Interval
-----------------------------------------------------------------------
      AGE            .02607      .00284     9.17      .02050      .03165
      HSAT         -.69985      .00908    77.06    -.71765    -.68205
   *  MARRIED        .01186      .06595      .18    -.11740      .14113
      EDUC         -.03530      .01403     2.52    -.06280    -.00781
      HHNINC       -.85642      .18247     4.69   -1.21406    -.49878
   *  HHKIDS       -.29073      .06661     4.36    -.42129    -.16018
-----------------------------------------------------------------------
```

What About Just Using LS?

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
<u>**Least Squares Coefficient Estimates**</u>					
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
<u>**Estimated Partial Effects**</u>					
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**	