Department of Economics

Econometric Analysis of Panel Data

Professor William Greene
Phone: 212.998.0876
Office: KMC 7-78
Home page:www.stern.nyu.edu/~wgreene
Email: wgreene@stern.nyu.edu
URL for course web page:
www.stern.nyu.edu/~wgreene/Econometrics/PanelDataEconometrics.htm

Assignment 5
Nonlinear Models

Part I. Weibull Regression Model

In class, we examined a 'loglinear,' exponential regression model,

$$f(y_i \mid \mathbf{x_i},1) = \frac{1}{\theta_i}\exp\left(-\frac{y_i}{\theta_i}\right),\ \theta_i = \exp(\mathbf{x_i}'\boldsymbol{\beta}) = E[y_i|\mathbf{x_i}]$$

The Weibull model is an extension of the exponential model which adds a shape parameter, γ;

$$f(y_i \mid \mathbf{x_i},\gamma) = \frac{\gamma y_i^{\gamma-1}}{\theta_i^{\gamma}}\exp\left(-\left[\frac{y_i}{\theta_i}\right]^{\gamma}\right) \quad E[y_i|x_i]=\Gamma[(\gamma+1)/2]\ \theta_i\ =\ .5^*sqr(\pi)\ \text{ if } \gamma = 2.$$

The exponential model results when $\gamma = 1$. (This distribution looks like, but is not the gamma distribution we discussed in class.) An interesting special case is the Rayleigh distribution, which has $\gamma = 2$. The resulting density is

$$f(y_i \mid \mathbf{x_i},2) = \frac{2y_i}{\theta_i^2}\exp\left(-\left[\frac{y_i}{\theta_i}\right]^2\right)$$

One of the interesting things about the Rayleigh distribution is that $E[y|\mathbf{x_i}] = .5\sqrt{\pi}\ \theta_i$ (compared to θ_i for the exponential. $.5\sqrt{\pi}$ is approximately equal to 0.866.) One difference is the variance. The variance of the exponential variable is θ_i^2. The variance of the Rayleigh variable is $[\Gamma(2) - \Gamma^2(1.5)]\theta_i^2$.

Since $\Gamma(t) = t-1!$ for integer t, $\Gamma(2) = 1$. When t = an integer + .5, we can use the recurrence $\Gamma(t) = (t-1)\Gamma(t-1)$ until we reach $\Gamma(.5)$ which equals $\sqrt{\pi}$. Combining terms, then, the variance of the Rayleigh variable is $[1-(.5\sqrt{\pi})^2]\theta_i^2 = 0.2146\theta_i^2$.

a. The parameters $\boldsymbol{\beta}$ in the Rayleigh model could be estimated either by nonlinear least squares or by maximum likelihood. Which would be more efficient? Explain.

The MLE is efficient among consistent and asymptotically normally distributed estimators of β. The MLE will surpass NLS because it uses information about the distribution while the NLS estimator only uses information about the form of the conditional mean function. Both estimators are consistent.

b. Form the log likelihood and derive the expressions for the first order conditions for maximizing the log likelihood for the Weibull model.

$$\log f(y_i \mid \mathbf{x}_i, \theta_i, \gamma) = \log\gamma + (\gamma-1)\log y_i - \gamma\log\theta_i - (y_i/\theta_i)^\gamma, \theta_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i)$$

$$\log L = \sum_{i=1}^N \log f(y_i \mid \mathbf{x}_i, \theta_i, \gamma)$$

$$\frac{\partial \log L}{\partial \theta_i} = \sum_{i=1}^N -\frac{\gamma}{\theta_i} - \gamma\left(\frac{y_i}{\theta_i}\right)^{\gamma-1}\left(\frac{-y_i}{\theta_i^2}\right)$$

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^N \frac{\partial \log L}{\partial \theta_i}\frac{\partial \theta_i}{\partial \boldsymbol{\beta}} = \sum_{i=1}^N \left(\frac{\partial \log L}{\partial \theta_i}\right)\theta_i \mathbf{x}_i = \sum_{i=1}^N \left(-\gamma\left[1-\left(\frac{y_i}{\theta_i}\right)^\gamma\right]\right)\mathbf{x}_i = \mathbf{0}$$

$$\frac{\partial \log L}{\partial \gamma} = \sum_{i=1}^N \left[\frac{1}{\gamma} + \log y_i - \log\theta_i - (y_i/\theta_i)^\gamma \log(y_i/\theta_i)\right]$$

$$= \sum_{i=1}^N \left[\frac{1}{\gamma} + \left(1-\left(\frac{y_i}{\theta_i}\right)^\gamma\right)\log(y_i/\theta_i)\right] = 0$$

c. How would you test the null hypothesis of the Rayleigh model (γ=2) against the more general null of the Weibull model (γ unrestricted)?

(1) likelihood ratio test, fitting the Weibull model without restriction ($\log L_u$) and the Rayleigh model with the restriction ($\log L_R$), then $\chi^2[1] = 2[\log L_U - \log L_R]$. Chi-squared test, one degree of freedom. (2) Wald test. For this restriction, just fit the Weibull model without restriction, then do a simple "t test" against the null hypothesis that $\gamma = 2$. t = (c – 2)/standard error.

d. How would you test the null hypothesis of the Rayleigh model (γ=2) against the alternative of the Exponential model ($\gamma = 1$)?

It's not possible to test the simple null against the simple alternative. (Sorry, a trick question, I suppose.) A Bayesian approach might suggest the "posterior odds ratio" P(Rayleigh|data)/P(Exponential|data). Note that this is not a "test" as such – one would not reject one model or the other on this basis, but only modify one's prior belief as to which model is more likely to be "correct."

e. Maximum likelihood estimates of the parameters of the three models based on the German health data discussed in class appear below. Carry out the test in part c. Which of the three do you think is the appropriate model given the results below.

(1) The LR statistic is 2(12033.5 – 11918.69) = 229.62. The critical value is 3.84 so the null hypothesis is rejected.
(2) The Wald statistic would be $[(2.12853619 - 2)/.00466881]^2 = 757.946$. Same conclusion

f. In the Rayleigh model, show how to obtain the three available estimators of the asymptotic covariance matrix of the MLE of $\boldsymbol{\beta}$. Remember, you are not estimating γ (it equals 2), and the expected value of y_i is still θ_i.

To do this, we will need the first derivatives of the log likelihood, the second derivatives and the expected derivatives for the Rayleigh model. Remember, the parameter γ is now fixed at 2. We have the first derivative with respect to β for the Weibull model above, so we can just insert $\gamma = 2$ for the Rayleigh model. So, the first derivatives are

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum\nolimits_{i=1}^{N} \left(-2\left[1-\left(\frac{y_i}{\theta_i}\right)^2\right]\right)\mathbf{x}_i = \sum\nolimits_{i=1}^{N} \mathbf{g}_i$$

The Berndt, Hall, Hall and Hausman estimator is just the inverse of the sum of squares,

$$BHHH = \sum\nolimits_{i=1}^{N} \mathbf{g}_i\mathbf{g}_i' = \sum\nolimits_{i=1}^{N} \left(-2\left[1-\left(\frac{y_i}{\theta_i}\right)^2\right]\right)^{\mathbf{2}} \mathbf{x}_i\mathbf{x}_i' = 4\sum\nolimits_{i=1}^{N} \left[1-\left(\frac{y_i}{\hat{\theta}_i}\right)^2\right]^2 \mathbf{x}_i\mathbf{x}_i'$$

To compute this, we would obtain the MLE of β and just replace θ_i in the above expression with the estimated values. To obtain the other estimates, we need the second derivatives.

$$\frac{\partial^2 \log L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = \sum\nolimits_{i=1}^{N} \left(-2\left[-2\left(\frac{y_i}{\theta_i}\right)\left(\frac{-y_i}{\theta_i^2}\right)\right]\right)\mathbf{x}_i(\theta_i\mathbf{x}_i')$$
$$= \sum\nolimits_{i=1}^{N} -4\left(\frac{y^2_i}{\theta_i^2}\right)\mathbf{x}_i\mathbf{x}_i'$$

(I differentiated first with respect to θ_i then θ_i with respect to β.) Then, the estimator of the asymptotic covariance matrix would be the negative of the inverse of this matrix, computed using the MLE of β to compute θ_i. This would be,

$$\text{Est.Asy.Var}[\hat{\beta}] = \left[\sum\nolimits_{i=1}^{N} 4\left(\frac{y^2_i}{\theta_i^2}\right)\mathbf{x}_i\mathbf{x}_i'\right]^{-1}$$

For the third estimator, we need the negative inverse of the expected second derivatives. To get the expected value of the second derivative, we need the expected value of y_i^2. The mean and variance of the Rayleigh variable are given in the problem. The expected value of y_i^2 is the variance plus the square of the mean, which, nicely, simplifies to just θ_i^2. Therefore, the negative expectatation of the second derivatives matrix is just

$-E[\partial^2 \log L/\partial\beta\partial\beta'] = 4\,\mathbf{X'X}$ which means that the estimator of the asymptotic covariance matrix is

$\text{Est.Asy.Var}[\hat{\beta}] = [4\mathbf{X'X}]^{-1}$!

```
+---------------------------------------------+
| Weibull (Loglinear) Regression Model        |
| Dependent variable              HHNINC      |
| Number of observations           27322      |
| Log likelihood function       12033.50      |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Parameters in conditional mean function
 Constant      3.44054643      .02266279   151.815   .0000
 EDUC          -.10914142      .00147212   -74.139   .0000    11.3201838
 MARRIED       -.31230818      .00750583   -41.609   .0000     .75869263
 AGE            .00053144      .00044049     1.206   .2276    43.5271942
          Shape parameter for Weibull model
 P_scale       2.12853619      .00466881   455.905   .0000

+---------------------------------------------+
| Exponential (Loglinear) Regression Model    |
| Log likelihood function       1539.191      |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Parameters in conditional mean function
 Constant      1.82555590      .04219675    43.263   .0000
 EDUC          -.05545277      .00267224   -20.751   .0000    11.3201838
 MARRIED       -.23664845      .01460746   -16.201   .0000     .75869263
 AGE            .00087436      .00057331     1.525   .1272    43.5271942

+---------------------------------------------+
| Weibull (Loglinear) Regression Model        |
| Log likelihood function       11918.69      |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Parameters in conditional mean function
 Constant      3.28524659      .02586426   127.019   .0000
 EDUC          -.10377049      .00172163   -60.275   .0000    11.3201838
 MARRIED       -.31371176      .00871996   -35.976   .0000     .75869263
 AGE            .00064343      .00048739     1.320   .1868    43.5271942
          Shape parameter for Weibull model
 P_scale       1.99999964   ......(Fixed Parameter).......
```

Part II. Marginal Effects in a Heteroscedastic Probit Model

Consider the following extension of the probit model. We make the disturbance heteroscedastic:

$$y_i^* = \alpha + x_{i1}\beta_1 + x_{i2}\beta_2 + \varepsilon_i$$
$$\varepsilon_i \sim N[0,\sigma_i^2] \text{ where } \sigma_i = \exp(\gamma_1 x_{i1} + \gamma_2 x_{i3})$$

This extension produces the probability model

$$\text{Prob}[y_i = 1 \mid x_{i1}, x_{i2}, x_{i3}] = \Phi\left(\frac{\alpha + x_{i1}\beta_1 + x_{i2}\beta_2}{\exp(x_{i1}\gamma_1 + x_{i3}\gamma_3)}\right)$$

Derive the partial (marginal) effects for this model, $\partial\text{Prob}(y_i=1)/\partial x_{i1}$, $\partial\text{Prob}(y_i=1)/\partial x_{i2}$, and $\partial\text{Prob}(y_i=1)/\partial x_{i3}$. It's worth noting that the partial effect for x_{i3} has the opposite sign from the coefficient.

Since x_{i1} appears in both numerator and denominator, we must differentiate the parts separately and add them.
$\partial\text{Prob}(y_i=1)/\partial x_{i1}$ =

$$\phi\left(\frac{x_{i1}\beta_1 + x_{i2}\beta_2}{\exp(x_{i1}\gamma_1 + x_{i3}\gamma_3)}\right)\times\left(\beta_1\frac{1}{\exp(x_{i1}\gamma_1 + x_{i3}\gamma_3)} + \frac{-x_{i1}\beta_1 + x_{i2}\beta_2}{[\exp(x_{i1}\gamma_1 + x_{i3}\gamma_3)]^2}\exp(x_{i1}\gamma_1 + x_{i3}\gamma_3)\gamma_1\right)$$
$$= \phi\left(\frac{x_{i1}\beta_1 + x_{i2}\beta_2}{\exp(x_{i1}\gamma_1 + x_{i3}\gamma_3)}\right)\times\left(\beta_1\frac{1}{\exp(x_{i1}\gamma_1 + x_{i3}\gamma_3)} - \frac{x_{i1}\beta_1 + x_{i2}\beta_2}{\exp(x_{i1}\gamma_1 + x_{i3}\gamma_3)}\gamma_1\right)$$

Essentially the same computation, but x_{i2} is only in the numerator, so the second term is zero.
$\partial\text{Prob}(y_i=1)/\partial x_{i2}$ =

$$\phi\left(\frac{x_{i1}\beta_1 + x_{i2}\beta_2}{\exp(x_{i1}\gamma_1 + x_{i3}\gamma_3)}\right)\times\left(\beta_2\frac{1}{\exp(x_{i1}\gamma_1 + x_{i3}\gamma_3)}\right)$$

Again, essentially the same, but now only the term in the denominator.

$$\phi\left(\frac{x_{i1}\beta_1 + x_{i2}\beta_2}{\exp(x_{i1}\gamma_1 + x_{i3}\gamma_3)}\right)\times\left(-\frac{x_{i1}\beta_1 + x_{i2}\beta_2}{\exp(x_{i1}\gamma_1 + x_{i3}\gamma_3)}\gamma_3\right)$$

It's worth noting that the partial effect for x_{i3} has the opposite sign from the coefficient.

Part III. Binomial Loglinear Model

Theory "Z" states that the age and education of the mother have an influence on the probability that a child will be female. Theory "Not Z" says that these two variables are irrelevant. Theory "There is no Theory" goes even further and states that the probability is always exactly one half. Consider modeling the number of female children, Girls_i in a sample of families; the number of children is Kids_i. The model in question is

$$\text{Kids}_i = \text{total number of children} = 0,1,\ldots$$

$$\text{Girls}_i = \text{number of female children} = 0,1,\ldots,K_i$$

$$\text{Prob}(\text{GIRLS} = \text{Girls}_i \mid \mathbf{x_i}, \text{Kids}_i) = \binom{\text{Kids}_i}{\text{Girls}_i} \theta_i^{\text{Girls}_i} (1-\theta_i)^{\text{Kids}_i - \text{Girls}_i}$$

$$0 < \theta_i < 1,\ \theta_i = \text{probability of a female child}$$

$$\theta_i = \frac{\exp(\mathbf{x_i'\beta})}{1+\exp(\mathbf{x_i'\beta})},\ \mathbf{x}_i = (1, \text{Age}_i, \text{Educ}_i), \beta = (\beta_0, \beta_1, \beta_2)$$

(Note that if $\text{Kids}_i = 0$, the probability that Girls_i equals zero is 1.).

The three theories are: Z = all three coefficients nonzero

Not Z $= \beta_1 = \beta_2 = 0$, β_0 unrestricted

No Theory $= \beta_0 = \beta_1 = \beta_2 = 0$

1. Derive the log likelihood for estimation of the three unknown parameters. (Note, the factorial term at the beginning of the probabilities does not involve the parameters, so it can be ignored. This is often labeled "an irrelevant constant."

$$\text{Prob}(\text{GIRLS} = \text{Girls}_i \mid \mathbf{x_i}, \text{Kids}_i) = \binom{\text{Kids}_i}{\text{Girls}_i} \theta_i^{\text{Girls}_i} (1-\theta_i)^{\text{Kids}_i - \text{Girls}_i}$$

$$\log L = \sum\nolimits_{i=1}^{N} \log \text{Prob}(\text{GIRLS} = \text{Girls}_i \mid \mathbf{x_i}, \text{Kids}_i)$$

$\text{Prob}[\text{Girls}_i|\mathbf{x}_i]$ =

$$= \sum\nolimits_{i=1}^{N} \log\binom{\text{Kids}_i}{\text{Girls}_i} + \text{Girls}_i \log\theta_i + (\text{Kids}_i - \text{Girls}_i)\log(1-\theta_i)$$

where θ_i is as defined earlier in terms of β.

2. Derive the first order conditions for maximizing your log likelihood function.

$$\log L = \sum\nolimits_{i=1}^{N} \log\binom{\text{Kids}_i}{\text{Girls}_i} + \text{Girls}_i \log\theta_i + (\text{Kids}_i - \text{Girls}_i)\log(1-\theta_i)$$

$$\frac{\partial \log L}{\partial \beta} = \sum\nolimits_{i=1}^{N} \frac{\partial \log L}{\partial \theta_i}\frac{\partial \theta_i}{\partial \beta} = \sum\nolimits_{i=1}^{N} \left[\frac{\text{Girls}_i}{\theta_i} - \frac{\text{Kids}_i - \text{Girls}_i}{1-\theta_i}\right]\theta_i(1-\theta_i)\mathbf{x}_i = \mathbf{0}$$

3. Discuss exactly how you will test the hypothesis of theory “Not Z” against the alternative of theory “Z.” How will you test the hypothesis of “No Theory” against theory “Z.” What statistics will you use.

To test the hypothesis of Theory Z against Not Theory Z, I would fit the model twice. Theory Z has all three elements of x_i in the model. Theory Not Z has only a constant term in the model. Then, I could use a likelihood ratio test to test the null hypothesis that the two coefficients are zero. Alternatively, I could use a Wald test based on the full model with all three nonzero coefficients.

4. The data you need to do your estimation and carry out your tests are placed in two formats on the course website, .xls for a spreadsheet and .csv is an ascii text file. The files contain 500 observations on Age, Educ, Kids, Girls. Use these data to estimate your model and test the hypotheses.

http://pages.stern.nyu.edu/~wgreene/Econometrics/BinomialData.xls
http://pages.stern.nyu.edu/~wgreene/Econometrics/BinomialData.csv

(Disclaimer: The data are completely synthetic – simulated with a random number generator. This is a numerical example, not a study based on actual outcomes.)
Tip: Once you have read the data into NLOGIT, you can compute your estimates with

```
maximize
; labels=beta0,beta1,beta2 ; start = 0,0,0
; fcn = bx = beta0+beta1*educ+beta2*age |
        ti = exp(bx)/(1+exp(bx)) |
        girls * log(ti) + (kids-girls)*log(1-ti) $
```

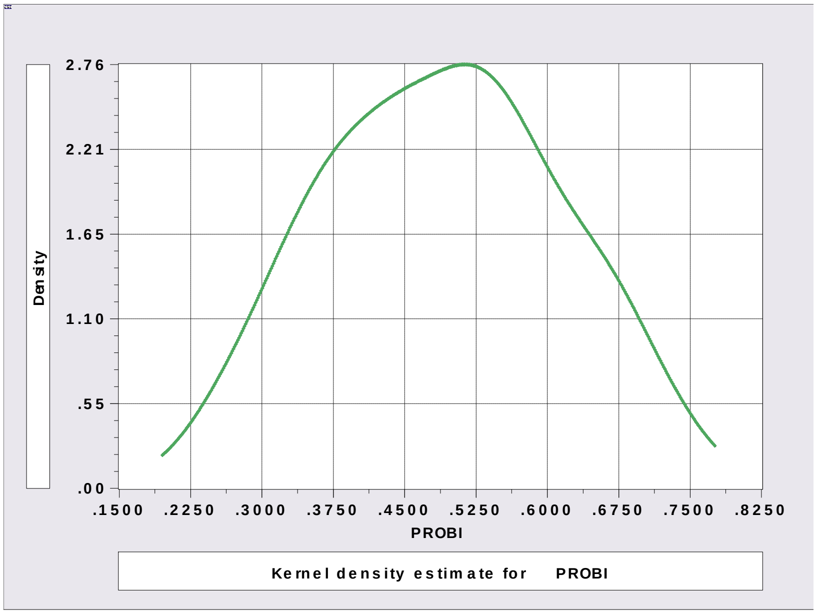
To fix certain coefficients to zero, one convenient way is to use ;FIX=list. For example, to force β_2 to equal zero in the results, you would add ;Fix=beta2 to the command. (This forces the estimate to equal the starting value(s).) Also, note that in your results, what NLOGIT reports as the “Log Likelihood” in its results is actually the negative of the log likelihood.

```
--> maximize
    ; labels=beta0,beta1,beta2 ; start = 0,0,0
    ; fcn = bx = beta0+beta1*educ+beta2*age |
    ti = exp(bx)/(1+exp(bx)) |
    girls * log(ti) + (kids-girls)*log(1-ti) $
Normal exit from iterations. Status=0. F=    1640.785

+---------------------------------------------+
| User Defined Optimization                   |
| Maximum Likelihood Estimates                |
| Dependent variable            Function      |
| Weighting variable                None      |
| Number of observations             500      |
| Iterations completed                 7      |
| Log likelihood function       1640.785      |
| Number of parameters                 0      |
| Info. Criterion: AIC =        -6.56314      |
|   Finite Sample: AIC =        -6.56314      |
| Info. Criterion: BIC =        -6.56314      |
| Info. Criterion:HQIC =        -6.56314      |
| Restricted log likelihood     .0000000      |
| Chi squared                   3281.570      |
| Degrees of freedom                   3      |
| Prob[ChiSqd > value] =        .0000000      |
| Model estimated: Apr 08, 2009, 05:41:02PM   |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]|
+--------+--------------+----------------+--------+--------+
|BETA0   |   -2.50069***       .28392633    -8.808   .0000 |
|BETA1   |     .05402***       .01498679     3.605   .0003 |
```

```
|BETA2   |     .03995***       .00377150    10.593   .0000 |
+--------+------------------------------------------------+
| Note: ***, **, * = Significance at 1%, 5%, 10% level.    |
+----------------------------------------------------------+

--> calc ; logl1 = logl $
--> matrix ; bx =b(2:3) ; vbx = varb(2:3,2:3) $
--> matrix ; list ; wald = bx'<vbx>bx $

Matrix WALD     has  1 rows and  1 columns.
               1
        +--------------
       1|  126.51465  **************** Hypothesis is rejected. Critical value
                           For chi squared with 2 degrees of freedom is 5.99
--> maximize
    ; labels=beta0,beta1,beta2 ; start = 0,0,0
    ; fcn = bx = beta0+beta1*educ+beta2*age |
    ti = exp(bx)/(1+exp(bx)) |
    girls * log(ti) + (kids-girls)*log(1-ti)
    ; fix = beta1,beta2$
NOTE: Convergence in initial iterations is rarely
at a true function optimum. This may not be a
solution (especially if initial iterations stopped).
Exit from iterative procedure.    3 iterations completed.
Check convergence values shown below.
Gradient value: Tolerance= .1000D-05, current value= .7260D-07
Function chg. : Tolerance= .0000D+00, current value= .1497D-08
Parameters chg: Tolerance= .0000D+00, current value= .1869D-05
Smallest abs. param. change from start value = .3884D-01
Normal exit from iterations. Status=0. F=    1712.994

+----------------------------------------------+
| User Defined Optimization                    |
| Maximum Likelihood Estimates                 |
| Dependent variable              Function     |
| Weighting variable                  None     |
| Number of observations               500     |
| Iterations completed                   3     |
| Log likelihood function         1712.994     |
| Number of parameters                   0     |
| Info. Criterion: AIC =          -6.85198     |
|   Finite Sample: AIC =          -6.85198     |
| Info. Criterion: BIC =          -6.85198     |
| Info. Criterion:HQIC =          -6.85198     |
| Restricted log likelihood       .0000000     |
| Chi squared                     3425.988     |
| Degrees of freedom                     1     |
| Prob[ChiSqd > value] =          .0000000     |
| Model estimated: Apr 08, 2009, 05:43:47PM    |
+----------------------------------------------+
+--------+-------------+----------------+--------+--------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]|
+--------+-------------+----------------+--------+--------+
|BETA0   |    -.03884          .03495948    -1.111   .2666 |
|BETA1   |       .000       ......(Fixed Parameter)....... |
|BETA2   |       .000       ......(Fixed Parameter)....... |
+--------+------------------------------------------------+
| Note: ***, **, * = Significance at 1%, 5%, 10% level.    |
+----------------------------------------------------------+
+-----------------------------------------------------------------------+
|Fixed Parameter... indicates a parameter that is constrained to equal|
|a fixed value (e.g., 0) or a serious estimation problem. If you did  |
|not impose a restriction on the parameter, check for previous errors.|
+-----------------------------------------------------------------------+

--> calc ; logl0 = logl $
--> calc ; list ; lrtest = 2*(logl1 - logl0) $
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 LRTEST  =    144.417353  Hypothesis is rejected again.
```

5. Using your results for for Theory Z, compute the probabilities that are predicted for the data set, and show the distribution with a kernel density estimator.

Create ; Probi = Lgp(b(1)+b(2)*educ+b(3)*age) $
Kernel ; Rhs = Probi $



Sort of interesting.

6. The expected number of Girls in a family with $Kids_i$ children is

$$E[Girls_i|Kids_i,x_i] = \theta_i \times Kids_i.$$

What is the partial effect with respect to Age? I.e., $\partial E[Girls_i|Kids_i,x_i]/\partial Age_i$ computed at the mean of age and education. Hint: θ_i, the probability, is the logit probability, $\Lambda(\beta'x)$. The derivative of $\Lambda(t)$ with respect to t is $d\Lambda(t)/dt = \Lambda(t)[1 - \Lambda(t)]$.

The partial effect would be

$Kids_i \times \partial \theta_i/\partial Age_i = \delta_1 = Kids_i \times \theta_i \times (1 - \theta_i) \times \beta_1$. Suppose this is computed at the means of the data. Just call this d_1. Denote θ_i computed at the means as just θ without a subscript. To compute an asymptotic variance for this, we need the Jacobian,

$$\partial d_1/\partial\beta = \overline{Kids}\ \{\{\theta(1-\theta)[0,1,0]\} + \beta_1(1-2\theta)[\theta(1-\theta)][1, \overline{Age}, \overline{Educ}]\}$$

.
Call this vector g_1'. Call the estimated asymptotic covariance matrix for the MLE V. Then, the estimator of the variance for d_1 is $g_1'Vg_1$.

```
calc ; kbar = xbr(kids)
     ; abar = xbr(age)
     ; ebar = xbr(educ) $
calc ; tbar = lgp(b(1)+b(2)*abar+b(3)*ebar)
     ; g1 = kbar*tbar*(1-tbar) ; g2 = b(2)*(1-2*tbar)*tbar*(1-tbar) $
matrix ; list ; j1= g1*[0,1,0]
    ; j2= g2*[1,abar,ebar] ; vd = j*varb*j'
    ; j = j1 + j2 $
    ; vd = j*varb*j'$
calc  ; list ; d1 = kbar * tbar*(1-tbar)*b(2) $
calc  ; list ; td1 = d1 / sqr(vd) $
Matrix J1       has  1 rows and  3 columns.
               1             2             3
        +------------------------------------------
       1|  .0000000D+00    1.20766    .0000000D+00

Matrix J2       has  1 rows and  3 columns.
               1             2             3
        +------------------------------------------
       1|    -.00200      -.08081      -.03106

Matrix VD       has  1 rows and  1 columns.
               1
        +--------------
       1|     .00030

Matrix J        has  1 rows and  3 columns.
               1             2             3
        +------------------------------------------
       1|    -.00200      1.12685      -.03106


--> calc  ; list ; d1 = kbar * tbar*(1-tbar)*b(2) $
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 D1      =       .065241
--> calc ; list ; td1 = d1 / sqr(vd) $
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 TD1     =      3.758577
```

Note that the statistical significance of this estimate is the same as that for the corresponding coefficient. Looking at the two parts of the Jacobian, it is easy to see why.

Part IV. Odds Ratio in the Logit Model

The results below present logit estimates of a model of whether the number of doctor visits is greater than zero based on the health care data discussed in class. (We used this example in class.)

```
+-------------------------------------------+
| Logit Model                               |
| Dependent variable               DOCTOR   |
| Number of observations            27326   |
| Log likelihood function       -17407.69   |
| Restricted log likelihood     -18016.64   |
| Chi squared                    1217.911   |
+-------------------------------------------+
+---------+-------------+---------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+-------------+---------------+--------+---------+----------+
          Characteristics in numerator of Prob[Y = 1]
 HHNINC        -.13813513      .07764383    -1.779   .0752      .35213516
 HHKIDS        -.25400914      .02984645    -8.511   .0000      .40271576
 EDUC          -.02375730      .00578666    -4.106   .0000    11.3201838
 MARRIED        .11799754      .03374477     3.497   .0005      .75869263
 AGE            .01811793      .00132457    13.678   .0000    43.5271942
 FEMALE         .53279823      .02817810    18.908   .0000      .47880829
 WORKING       -.15388095      .03185320    -4.831   .0000      .67714662
 Constant      -.05351200      .09905516     -.540   .5890
```

a. The results given are estimates of the coefficients, β. Researchers are sometimes interested in "odds" ratios, which are computed as $\exp(\beta)$. (See, for example, the Stata manual, volume 2, G-M.) How would the results in the table above change if we reported these, instead? Show explicitly.

The "coefficients" reported would be $\exp(\beta)$ instead, so, for example, the -.13813513 would be reported as exp(-.13813513) = 0.870981. To compute the standard errors, we would use the delta method. The variance of exp(b) is $[\exp(b)]^2\text{Var}[b]$, which means that the standard errors would be multiplied by the square root, which is exp(b). Thus, the first reported standard error would become (.870981).07764383 = .0676263. The "t-ratios" would be computed differently. The test of whether the "odds ratio" equals zero would make no sense, since $\exp(\beta)$ cannot equal zero. The interesting hypothesis is whether $\exp(\beta) = 1$. The test statistic would be

$$[\exp(b) - 1] / \text{New Standard error} = [\exp(b) - 1]/[\exp(b)\times\text{Old standard error}]$$
$$= [\exp(b) - 1] / [\exp(b)\times s_b] = (1/s_b) - 1/[\exp(b)\times s_b]$$

For the first one, this would be (0.870981 – 1)/.0676263 = -1.908.

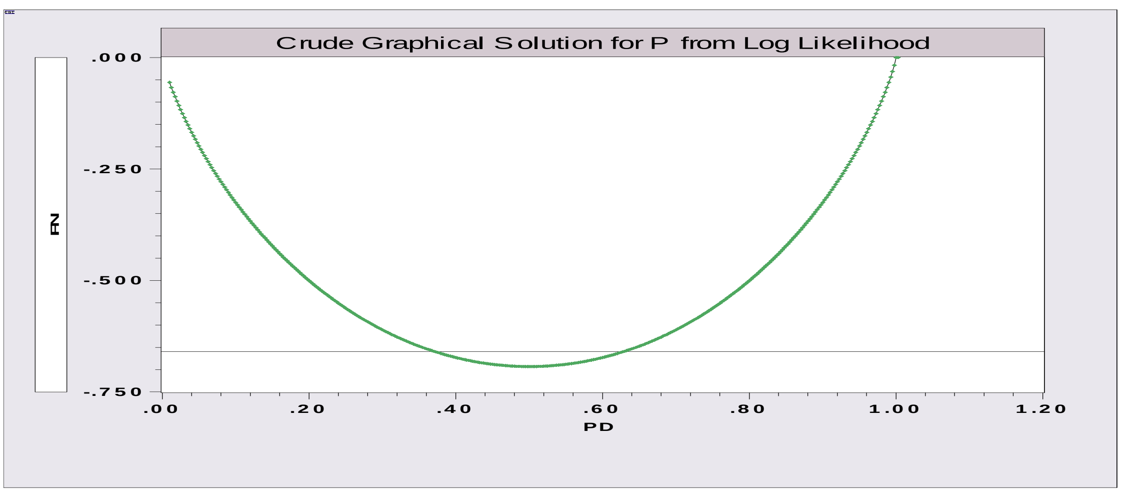
b. The restricted log likelihood in a binary choice model is computed for a model which contains only a constant term. This, in turn, ultimately is a function of the proportion of ones in the sample. Given the value above, deduce the number of observations for which DOCTOR equals 1 in the sample of 27,326. (Hint: there are two solutions – the problem is symmetric in P and (1-P). The correct solution is the larger one.)

In the model with only a constant term, logL = N[PlogP + (1-P)log(1-P] = -18016.64. This means that [PlogP + (1-P)log(1-P] = -18016.64/27326 = -0.6592988.
There are a few ways to approach this: One is to solve the problem numerically:

```
sample ; 1 $
mini ; start = .8
     ; labels = p
     ; fcn =  ((p*log(p)+(1-p)*log(1-p)) + .6592988) ^2 $
+---------------------------------------------+
| User Defined Optimization                   |
| Maximum Likelihood Estimates                |
| Dependent variable              Function    |
| Weighting variable                  None    |
| Number of observations                 1    |
| Iterations completed                   5    |
| Log likelihood function       .8399616E-16 |
| Number of parameters                   0    |
| Info. Criterion: AIC =             .00000   |
|   Finite Sample: AIC =             .00000   |
| Info. Criterion: BIC =             .00000   |
| Info. Criterion:HQIC =             .00000   |
| Restricted log likelihood       .0000000    |
| Chi squared                   .1679923E-15 |
| Degrees of freedom                     1    |
| Prob[ChiSqd > value] =          1.000000    |
| Model estimated: Apr 08, 2009, 08:30:19PM   |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]|
+--------+--------------+----------------+--------+--------+
|P       |     .62935        1.00000000      .629    .5291 |
+--------+-------------------------------------------------+
```

If you use .2 as the starting value, you get the other solution, .37065. Another approach is to plot the function.

```
sample;1-395$
create ; pd = trn(.01,.0025)$
create ; fn = pd*log(pd)+(1-pd)*log(1-pd)$
plot   ; lhs = pd ; rhs = fn ; bars = -.6592988
       ; fill
       ; endpoints = 0,1
;title=Crude Graphical Solution for P from Log Likelihood$
```



If P = .62935, then the number of ones in the sample is .62935(27326) = 17,196. (The actual value is 17,191. The difference is due to rounding error in the solution.)

Part V. The Poisson Regression Model

The following is based on the health care data used in several previous examples. We consider fitting a Poisson regression model to the variable DOCVIS which is the number of visits to the doctor by the individual in the given period. The model is as follows:

$$\text{Prob}[\text{DocVis}_i = y_i \mid \mathbf{x}_i] = \frac{\exp(-\theta_i)\theta_i^{y_i}}{y_i!}, y_i = 0,1,\ldots,\ \theta_i = \exp(\mathbf{x}_i'\boldsymbol{\beta})$$

a. Derive the log likelihood function for estimating $\boldsymbol{\beta}$ from a sample of n observations on y_i and $\mathbf{x}_i$.

The log likelihood is the sum of the logs of the probabilities for the observed variable:

$$\log L = \sum_{i=1}^{N} -\theta_i + y_i \log\theta_i - \log(y_i!), \theta_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i)$$

b. This is yet another log linear model in which $E[y_i] = \theta_i$. Use this result to show that the expected values of the first derivatives of the log likelihood function have expectation zero.

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{N} \frac{\partial \log L}{\partial \theta_i}\frac{\partial \theta_i}{\partial \beta} = \sum_{i=1}^{N}\left[-1 + \frac{y_i}{\theta_i}\right]\theta_i\mathbf{x}_i = \sum_{i=1}^{N}[y_i - \theta_i]\mathbf{x}_i$$

Since $E[y_i|x_i] = \theta_i$, each term obviously has expectation zero, so the sum does. Notice that the first order condition is $\Sigma_i\, e_i\mathbf{x}_i = \mathbf{0}$ where e_i is a residual. This is equivalent to $X'e = 0$, which is familiar.

c. Derive the forms of the three estimators of the asymptotic covariance matrix.

The BHHH estimator is just the inverse of the sum of squares of the first derivatives,

$$\text{BHHH} = \left[\sum_{i=1}^{N}[y_i - \theta_i]^2\mathbf{x}_i\mathbf{x}_i'\right]^{-1}$$

The second derivatives matrix is simple, since $\partial\theta_i/\partial\beta = \theta_i x_i$. Therefore,

$$\frac{\partial^2 \log L}{\partial\beta\partial\beta'} = \sum_{i=1}^{N} -\theta_i\mathbf{x}_i\mathbf{x}_i'$$

The estimated asymptotic covariance matrix is the negative inverse of this matrix,

$$\text{Est.Asy.Var} = \left[\sum_{i=1}^{N}\theta_i\mathbf{x}_i\mathbf{x}_i'\right]^{-1}$$

This is computed just by inserting the MLE in the expression. Since this function does not involve y_i, this is also the expectation. The third form is the same as the second.

d. Show that the restricted log likelihood in which $\mathbf{x}_i$ contains only a constant term is a function only of the sample mean of y_is.

If there is only a constant term, then $\theta_i = \exp(\beta_0)$ and

$$\log L = \sum_{i=1}^{N} -\theta_i + y_i \log\theta_i - \log(y_i!), \theta_i = \exp(\beta_0)$$

$$\log L = N\exp(\beta_0) + \sum_{i=1}^{N} \beta_0 y_i - \sum_{i=1}^{N} \log(y_i!)$$

$$= N[\exp(\beta_0) + \beta_0 \bar{y}] \ + \text{ the irrelevant constant.}$$

e. Using the health care data set, estimate a Poisson model for DOCVIS in which

$\mathbf{x}_i$=[1, `female,age,hhninc,hhkids,educ,married]`.

```
+---------------------------------------------+
| Poisson Regression                          |
| Maximum Likelihood Estimates                |
| Model estimated: Apr 08, 2009 at 08:57:55PM.|
| Dependent variable               DOCVIS     |
| Weighting variable                 None     |
| Number of observations            27326     |
| Iterations completed                  7     |
| Log likelihood function       -103727.3     |
| Number of parameters                  7     |
| Info. Criterion: AIC =          7.59235     |
|   Finite Sample: AIC =          7.59235     |
| Info. Criterion: BIC =          7.59446     |
| Info. Criterion:HQIC =          7.59303     |
| Restricted log likelihood     -108662.1     |
| McFadden Pseudo R-squared      .0454145     |
| Chi squared                    9869.679     |
| Degrees of freedom                    6     |
| Prob[ChiSqd > value] =         .0000000     |
+---------------------------------------------+
+---------------------------------------------+
| Poisson Regression                          |
| Chi- squared =255127.59573  RsqP=   .0818   |
| G  - squared =154416.01169  RsqD=   .0601   |
| Overdispersion tests: g=mu(i)  : 20.974     |
| Overdispersion tests: g=mu(i)^2: 20.943     |
+---------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	.77266707	.02813535	27.463	.0000	
FEMALE	.29287271	.00701806	41.731	.0000	.47877479
AGE	.01763160	.00034644	50.894	.0000	43.5256898
HHNINC	-.52228656	.02258946	-23.121	.0000	.35208362
HHKIDS	-.16031757	.00840186	-19.081	.0000	.40273000
EDUC	-.02981125	.00174594	-17.075	.0000	11.3206310
MARRIED	.00964101	.00874426	1.103	.2702	.75861817

f. Using your estimator, test the hypothesis that all coefficients in the model except the constant term are zero. The easiest test to use will be the likelihood ratio test. Show how to do the Lagrange multiplier test. (It has a particularly simple form in this model.) If you have access to the necessary matrix computations, carry out the LM test.

```
A Wald test of the hypothesis:
--> Matrix ; b1 = b(2:7) ; v1 = varb(2:7,2:7)$
--> Matrix ; list ; Wald = b1'<v1>b1 $
Matrix WALD     has  1 rows and  1 columns.
               1
        +--------------
       1| 9873.01682
--> Calc   ; list ; ctb(.95,6)$
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 Result  =     12.591587
```

```
Likelihood ratio test.
--> Calc ; l1 = logl $
--> Poisson ; Lhs = DocVis ; Rhs = one $
--> Calc ; l0 = logl $
--> Calc ; List ; LRTest = 2*(l1 - l0) $
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 LRTEST  =   9869.679159
```

For the Lagrange Multiplier test, you are going to compute the first derivatives,

$\partial\log L/\partial\beta \;=\; \sum_{i=1}^{N}\left[y_i-\theta_i\right]\mathbf{x}_i$ at the restricted estimates in which all the coefficients except the constant term are zero. Call this

$$\mathbf{g}_0=\sum_{i=1}^{N}\left[y_i-\theta_i\right]\mathbf{x}_i=\sum_{i=1}^{N}\mathbf{g}_{i0}$$

You need the covariance matrix for the first derivatives. The easiest way to compute them is to use the sum of squares as usual. Then, use a Wald statistic to test the hypothesis that g_0 equals zero. Thus,

$$V_0 \;=\; \sum_{i=1}^{N}\mathbf{g}_{i0}\mathbf{g}_{i0}'=\sum_{i=1}^{N}(y_i-\theta_0)^2\,\mathbf{x}_i\mathbf{x}_i'$$

Collecting terms, the LM statistic will be $g_0'V_0^{-1}g_0$. I.e.,

$$\text{LM} \;=\; \left(\sum_{i=1}^{N}\left[y_i-\theta_0\right]\mathbf{x}_i\right)'\left(\sum_{i=1}^{N}\left[y_i-\theta_0\right]^2\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\left(\sum_{i=1}^{N}\left[y_i-\theta_0\right]\mathbf{x}_i\right)$$

```
--> Create  ; theta0 = exp(b(1)) $
--> Create  ; e0 = DocVis - theta0 ; e02 = e0*e0 $
--> Matrix  ; List ; LM = e0'X * <X'[e02]X> * X'e0 $
Matrix LM       has  1 rows and  1 columns.
               1
        +--------------
       1| 1110.82997
```

```
NLOGIT provides a way to automate LM statistics:

--> Poisson ; Lhs = DocVis ; Rhs = X
    ; Start = b(1),0,0,0,0,0,0 ; Maxit=0 $
+---------------------------------------------+
| Poisson Regression                          |
| Maximum Likelihood Estimates                |
| Model estimated: Apr 08, 2009 at 09:13:36PM.|
| Dependent variable               DOCVIS     |
| Weighting variable                 None     |
| Number of observations            27326     |
| Iterations completed                  2     |
| LM Stat. at start values       10030.87     | <========================
| LM statistic kept as scalar    LMSTAT       |
| Log likelihood function       -108662.1     |
| Number of parameters                  7     |
| Info. Criterion: AIC =          7.95353     |
|   Finite Sample: AIC =          7.95353     |
| Info. Criterion: BIC =          7.95564     |
| Info. Criterion:HQIC =          7.95421     |
+---------------------------------------------+

+---------------------------------------------+
| Poisson Regression                          |
| Chi- squared =277862.01305  RsqP=   .0000   |
| G  - squared =164285.69085  RsqD=   .0000   |
| Overdispersion tests: g=mu(i)  : 21.669     |
| Overdispersion tests: g=mu(i)^2: 21.669     |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    1.15798903       .02551994    45.376   .0000
 FEMALE  |       .000000       .00692075      .000  1.0000    .47877479
 AGE     |       .000000       .00034270      .000  1.0000   43.5256898
 HHNINC  |       .000000       .02028930      .000  1.0000    .35208362
 HHKIDS  |       .000000       .00799415      .000  1.0000    .40273000
 EDUC    |       .000000       .00156684      .000  1.0000   11.3206310
 MARRIED |       .000000       .00891090      .000  1.0000    .75861817
```

Notice the statistic is different from what we obtained above. The reason is that this procedure uses the second derivatives instead of the squares of the first derivatives to compute the covariance matrix. If you use

```
--> Matrix  ; List ; LM = e0'X * <X'[theta0]X> * X'e0 $
```

in the program above, you will reproduce the LM statistic above.

Estimating the Poisson Model.

All programs that you might use these days, Stata, SAS, SPSS, NLOGIT, EViews, have a pushbutton estimator for the Poisson model. But, this one, like the probit or logit models, is exceedingly simple to estimate, and you can program Newton's method and see how it works close up. The following shows how you can do this with NLOGIT. The annotations show what each command does. You should just put these commands on your editing screen, and execute them as shown below. (The lines with leading question marks are comments that can be ignored.) Based on part III, you should also be able to write a MAXIMIZE command to do the estimation. You might try this as well.

```
? (1) You have to load the Healthcare.lpj data set. I assume this is
? done.  The next line defines the variables in the equation as
? specified in the assignment.  Note, though that this also defines a
? matrix named X
     namelist ; x=one,female,age,hhninc,hhkids,educ,married$
? This next line shows you what you will be doing with your program.
? It fits the Poisson model using the internal estimator. We will
? replicate these results
     poisson ; lhs=docvis;rhs=x$
? Now, we obtain starting values for the iterations.  If all the slopes
? were zero, then E[y] would equal exp(α), so we can estimate the
? constant term with the log of the mean of the dependent variable.
? Then start the other coefficients at zero. The matrix command defines
? a column vector of this form.
     calc   ; list ; a0=log(xbr(docvis))$
     matrix ; beta = [a0/0/0/0/0/0/0] $
? This small set of commands does the iterations.  Note, the function
? involves the log of yi!.  We use Gamma(y+1) = y! and a special version,
? the log of the gamma function, lgm(y+1) = logy!
?*************************************************************************
? To do the iterations, highlight and execute these commands. When done,
? the calc command shows you g'H-1g. Execute the commands several times.
? You will see this go toward zero very quickly. When it gets very small,
? you are done iterating.  Then just display the results. Did you replicate
? the "real" results above?
     procedure $
     create ; ey = exp(beta'x)                 ? Mean
            ; logli = -ey + docvis*log(ey)     ? logL(i)
                 - lgm(docvis+1)               ? logL(i)
            ; gi = docvis - ey                 ? first derivative
            ; hi = ey $                        ? second derivative
? Matrix manipulations do the update of Newton's method.
     matrix ; score = X'gi
            ; Hessian  = X'[hi]X
            ; update = <Hessian>*score
            ; beta = beta + update $
     calc   ; list ; ghg = score'update $      ?
     endproc$
     execute ; n = 5 $
? Display results
     matrix ; stat(beta,<Hessian>,x)$
```

```
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 GHG     =  10030.869712
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 GHG     =    316.917939
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 GHG     =       .339091
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 GHG     =       .000001
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 GHG     =       .000000

Maximum repetitions of PROC

--> matrix ; stat(beta,<Hessian>,x)$

+---------------------------------------------------+
|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|     .77266707       .02813535    27.463   .0000
 FEMALE  |     .29287271       .00701806    41.731   .0000
 AGE     |     .01763160       .00034644    50.894   .0000
 HHNINC  |    -.52228656       .02258946   -23.121   .0000
 HHKIDS  |    -.16031757       .00840186   -19.081   .0000
 EDUC    |    -.02981125       .00174594   -17.075   .0000
 MARRIED |     .00964101       .00874426     1.103   .2702
```

```
calc   ; list ; a0=log(xbr(docvis))$
matrix ; beta = [a0/0/0/0/0/0/0] $
Maximize ; start = beta
         ; Labels = c0,c1,c2,c3,c4,c5,c6
         ; Fcn = ti = exp(c0'x) |
                 -ti + Docvis*log(ti) - lgm(Docvis+1) $
```

```
+----------------------------------------------+
| User Defined Optimization                    |
| Maximum Likelihood Estimates                 |
| Model estimated: Apr 08, 2009 at 09:28:58PM.|
| Dependent variable             Function      |
| Weighting variable                 None      |
| Number of observations            27326      |
| Iterations completed                 12      |
| Log likelihood function        103727.3      |
| Number of parameters                  0      |
| Info. Criterion: AIC =         -7.59184      |
|   Finite Sample: AIC =         -7.59184      |
| Info. Criterion: BIC =         -7.59184      |
| Info. Criterion:HQIC =         -7.59184      |
| Restricted log likelihood      .0000000      |
| Chi squared                    207454.6      |
| Degrees of freedom                    7      |
| Prob[ChiSqd > value] =         .0000000      |
+----------------------------------------------+
+--------+--------------+----------------+--------+--------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]|
+--------+--------------+----------------+--------+--------+
 C0      |     .77266698       .00942052    82.020   .0000
 C1      |     .29287272       .00220396   132.885   .0000
 C2      |     .01763161       .00011478   153.606   .0000
 C3      |    -.52228638       .00775505   -67.348   .0000
 C4      |    -.16031757       .00270115   -59.352   .0000
 C5      |    -.02981125       .00063915   -46.642   .0000
 C6      |     .00964094       .00267079     3.610   .0003
```