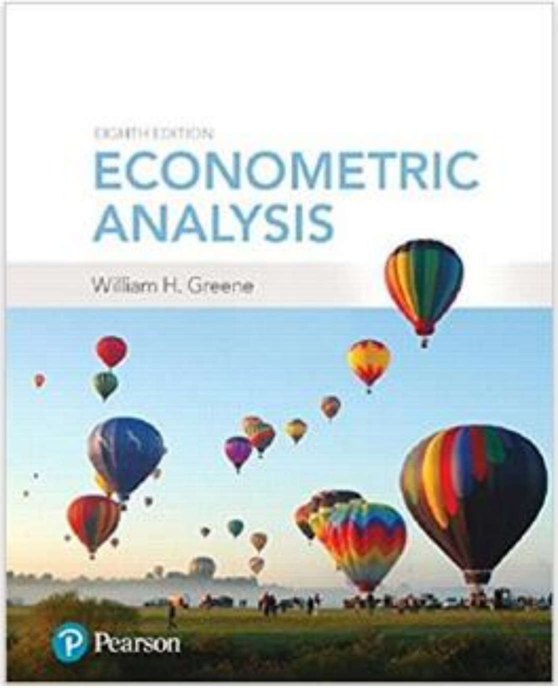
Econometrics I

Professor William Greene
Stern School of Business
Department of Economics



Econometrics I

Part 20 –MLE Applications and a Two Step Estimator

Poisson Regression Model

Application of ML Estimation: Poisson Regression for a Count of Events

Poisson Probability: $\text{Prob}[y=j] = \dfrac{\exp(-\lambda)\lambda^{j}}{j!},\ j = 0,1,\ldots$

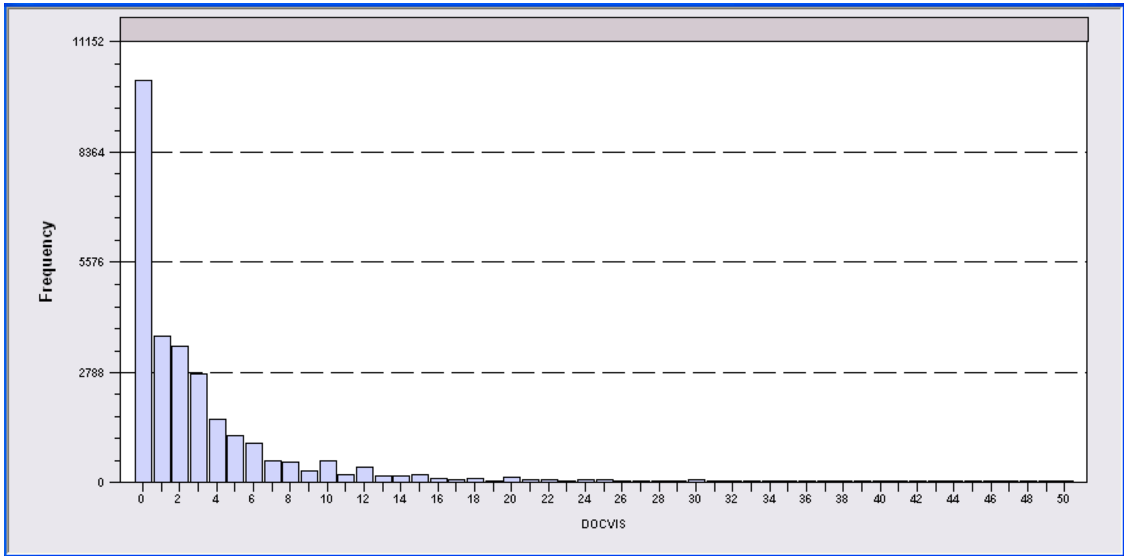
Regression Model: $\lambda = E[y|x] = \exp(\beta'x)$

Competing estimators: Nonlinear Least Squares - Consistent

Maximum Likelihood - Consistent and Efficient

Application: Doctor Visits

- German Individual Health Care data: n=27,236
- Model for number of visits to the doctor:
  - Poisson regression (fit by maximum likelihood)
  - Income, Education, Gender



Poisson Model

Density of Observed y

$$\text{Prob}[\text{y}_i = \text{j} \mid \mathbf{x}_i] = \frac{\exp(-\lambda_i)\lambda_i^j}{j!}$$

Log Likelihood

$$\text{logL}(\boldsymbol{\beta}|\mathbf{y},\mathbf{X}) = \sum_{i=1}^{n} \left[-\lambda_i + y_i \log\lambda_i - \log y_i!\right]$$

Likelihood Equations = Derivatives of log likelihood

$$\frac{\partial\lambda_i}{\partial\boldsymbol{\beta}} = \frac{\partial\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{\partial\boldsymbol{\beta}} = \exp(\boldsymbol{\beta}'\mathbf{x}_i)\frac{\partial\boldsymbol{\beta}'\mathbf{x}_i}{\partial\boldsymbol{\beta}} = \lambda_i\mathbf{x}_i$$

$$\frac{\partial\log L}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{n}\left[-\lambda_i\mathbf{x}_i + y_i\mathbf{x}_i\right] = \mathbf{0}$$

$$= \sum_{i=1}^{n}\mathbf{x}_i\left[y_i - \lambda_i\right] = \sum_{i=1}^{n}\mathbf{x}_i\varepsilon_i$$

Asymptotic Variance of the MLE

Variance of the first derivative vector:

Observations are independent. First derivative vector is the sum of n independent terms. The variance is the sum of the variances. The variance of each term is

$$\text{Var}[\mathbf{x}_i(y_i - \lambda_i)] = \mathbf{x}_i\mathbf{x}_i' \ \text{Var}[y_i - \lambda_i] = \lambda_i\mathbf{x}_i\mathbf{x}_i'$$

Summing terms

$$\text{Var}\left[\frac{\partial \log L}{\partial \beta}\right] = \sum_{i=1}^{n} \lambda_i \mathbf{x}_i\mathbf{x}_i' = \mathbf{X}'\mathbf{\Lambda}\mathbf{X}$$

Estimators of the Asymptotic Covariance Asymptotic Covariance Matrix

Conventional Estimator - Inverse of the information matrix

"Usual"

$$\left[\sum_{i=1}^{n}\lambda_i \mathbf{x}_i \mathbf{x}_i'\right]^{-1} = \left(\mathbf{X}'\mathbf{\Lambda}\mathbf{X}\right)^{-1}$$

"Berndt, Hall, Hall, Hausman" (BHHH)

$$\left[\sum_{i=1}^{n}\mathbf{g}_i \mathbf{g}_i'\right]^{-1} = \left[\sum_{i=1}^{n}[(y_i - \lambda_i)\mathbf{x}_i][(y_i - \lambda_i)\mathbf{x}_i']\right]^{-1}$$

$$= \left[\sum_{i=1}^{n}(y_i - \lambda_i)^2 \mathbf{x}_i \mathbf{x}_i'\right]^{-1}$$

Robust Estimation

- Robust (Sandwich) Estimator
- **$\mathbf{H}^{-1}$ ($\mathbf{G}'\mathbf{G}$) $\mathbf{H}^{-1}$**

$$\left(\mathbf{X}'\mathbf{\Lambda}\mathbf{X}\right)^{-1}\left[\sum_{i=1}^{n}(y_i-\lambda_i)^2\mathbf{x}_i\mathbf{x}_i'\right]\left(\mathbf{X}'\mathbf{\Lambda}\mathbf{X}\right)^{-1}$$

Partial Effects

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

To use the delta method, we need

$$\frac{\partial\boldsymbol{\delta}(\mathbf{x})}{\partial\boldsymbol{\beta}'}=\lambda(\mathbf{x})\mathbf{I}+\boldsymbol{\beta}\lambda(\mathbf{x})\mathbf{x}'=\mathbf{D}$$

$$\text{Est.Asy.Var}\left[\hat{\boldsymbol{\delta}}(\mathbf{x})\right]=\hat{\mathbf{D}}\,\text{Est.Asy.Var}\left[\hat{\boldsymbol{\beta}}\right]\hat{\mathbf{D}}'$$

Regression and Partial Effects

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	1.72492985	.02000568	86.222	.0000	
FEMALE	.31954440	.00696870	45.854	.0000	.47877479
HHNINC	-.52475878	.02197021	-23.885	.0000	.35208362
EDUC	-.04986696	.00172872	-28.846	.0000	11.3206310

```
+-----------------------------------------------+
| Partial derivatives of expected val. with     |
| respect to the vector of characteristics.     |
| Effects are averaged over individuals.        |
| Observations used for means are All Obs.      |
| Conditional Mean at Sample Point     3.1835   |
| Scale Factor for Marginal Effects    3.1835   |
+-----------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
FEMALE	1.01727755	.02427607	41.905	.0000	.47877479
HHNINC	-1.67058263	.07312900	-22.844	.0000	.35208362
EDUC	-.15875271	.00579668	-27.387	.0000	11.3206310

Comparison of Standard Errors

Negative Inverse of Second Derivatives

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	1.72492985	.02000568	86.222	.0000	
FEMALE	.31954440	.00696870	45.854	.0000	.47877479
HHNINC	-.52475878	.02197021	-23.885	.0000	.35208362
EDUC	-.04986696	.00172872	-28.846	.0000	11.3206310

BHHH

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]
Constant	1.72492985	.00677787	254.495	.0000
FEMALE	.31954440	.00217499	146.918	.0000
HHNINC	-.52475878	.00733328	-71.559	.0000
EDUC	-.04986696	.00062283	-80.065	.0000

Why are they so different? Model failure. This is a panel. There is autocorrelation.

MLE vs. Nonlinear LS

```
Poisson Regression
Dependent variable               DOCVIS
Log likelihood function    -90877.90223
Restricted log likelihood -108662.13583
Chi squared [   6 d.f.]     35568.46720
Significance level               .00000
McFadden Pseudo R-squared      .1636654
Estimation based on N =   27326, K =   7
Inf.Cr.AIC  = 181769.8 AIC/N =    6.652
Model estimated: Aug 25, 2011, 15:29:49
Chi- squared =188132.60552  RsqP= .3229
G  - squared =128717.22366  RsqD= .2165
Overdispersion tests: g=mu(i)  : 24.268
Overdispersion tests: g=mu(i)^2: 25.247
```

DOCVIS	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
Constant	2.54498***	.02800	90.91	.0000	2.49011	2.59985
AGE	.00784***	.00035	22.11	.0000	.00714	.00853
HSAT	-.22783***	.00133	-171.29	.0000	-.23044	-.22522
MARRIED	.00677	.00876	.77	.4393	-.01039	.02393
EDUC	-.02033***	.00171	-11.91	.0000	-.02368	-.01699
HHNINC	-.26664***	.02209	-12.07	.0000	-.30993	-.22335
HHKIDS	-.12510***	.00839	-14.91	.0000	-.14154	-.10865

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

Exponential Regression Model

$$P(y_i \mid \mathbf{x}_i) = \frac{1}{\theta_i}\exp(-y_i/\theta_i),$$

$$\theta_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i) = E[y_i \mid \mathbf{x}_i]; Var[y_i \mid \mathbf{x}_i] = \theta_i^2$$

$$LogL = \log\prod_{i=1}^{n} P(y_i \mid \mathbf{x}_i) = \sum_{i=1}^{n} -\log\theta_i - \frac{y_i}{\theta_i}$$

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \frac{\partial \log L_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n}\left(\frac{-1}{\theta_i} + \frac{y_i}{\theta_i^2}\right)\theta_i \mathbf{x}_i = \sum_{i=1}^{n}\left(\frac{y_i}{\theta_i} - 1\right)\mathbf{x}_i$$

$$\text{Note since } \theta_i = E[y_i \mid \mathbf{x}_i],\ \mathrm{E}\left[\frac{\partial \log L}{\partial \boldsymbol{\beta}}\right] = \mathbf{0}$$

Variance of the First Derivative

$$P(y_i \mid \mathbf{x}_i) = \frac{1}{\theta_i}\exp(-y_i/\theta_i),$$

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

Note since $\theta_i = E[y_i \mid \mathbf{x}_i]$, $\text{E}\left[\frac{\partial \log L}{\partial \boldsymbol{\beta}}\right] = \mathbf{0}$

$$\text{Var}\left[\frac{\partial \log L}{\partial \boldsymbol{\beta}}\right] = \sum_{i=1}^{n}\left(\frac{1}{\theta_i^2} Var[y_i \mid \mathbf{x}_i]\right)\mathbf{x}_i\mathbf{x}_i' = \sum_{i=1}^{n}\left(\frac{1}{\theta_i^2}\theta_i^2\right)\mathbf{x}_i\mathbf{x}_i' = \mathbf{X}'\mathbf{X}$$

Hessian

$$P(y_i \mid \mathbf{x}_i) = \frac{1}{\theta_i}\exp(-y_i/\theta_i),$$

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

$$\frac{\partial^2 \log L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = \sum_{i=1}^{n}\left(-\frac{y_i}{\theta_i^2}\right)\theta_i\,\mathbf{x}_i\mathbf{x}_i' = -\sum_{i=1}^{n}\left(\frac{y_i}{\theta_i}\right)\mathbf{x}_i\mathbf{x}_i'$$

$$-E\left[\frac{\partial^2 \log L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}\right] = \mathbf{X}'\mathbf{X}, \text{ because } \mathrm{E}[y_i \mid \mathbf{x}_i] = \theta_i$$

Variance Estimators

Negative inverse of actual second derivatives Matrix

$$\left[\sum_{i=1}^{n}\left(\frac{y_i}{\hat{\theta}_i}\right)\mathbf{x}_i\mathbf{x}_i'\right]^{-1}, \hat{\theta}_i = \exp\left[\hat{\boldsymbol{\beta}}'_{MLE}\mathbf{x}_i\right]$$

Negative inverse of expected second derivatives

$$-E\left[\frac{\partial^2 \log L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right] = \mathbf{X}'\mathbf{X}, \textbf{so } [\mathbf{X}'\mathbf{X}]^{-1}$$

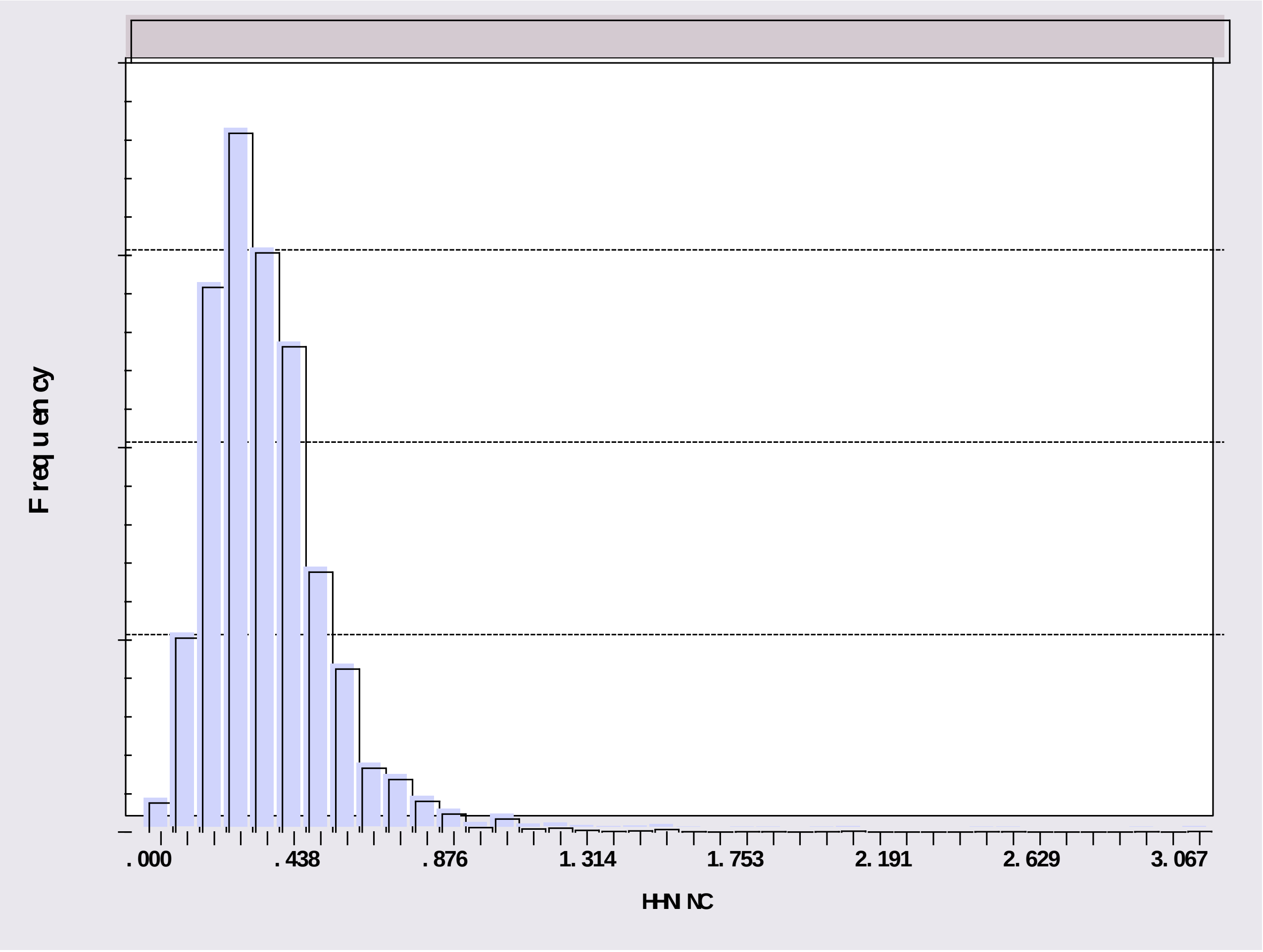
Sum of outer products of first derivatives (BHHH)

$$\left[\sum_{i=1}^{n}\left(\frac{y_i}{\hat{\theta}_i}-1\right)^{\mathbf{2}}\mathbf{x}_i\mathbf{x}_i'\right]^{-1}$$

"Robust" estimator in wide use

$$\left[\sum_{i=1}^{n}\left(\frac{y_i}{\hat{\theta}_i}\right)\mathbf{x}_i\mathbf{x}_i'\right]^{-1}\left[\sum_{i=1}^{n}\left(\frac{y_i}{\hat{\theta}_i}-1\right)^{\mathbf{2}}\mathbf{x}_i\mathbf{x}_i'\right]\left[\sum_{i=1}^{n}\left(\frac{y_i}{\hat{\theta}_i}\right)\mathbf{x}_i\mathbf{x}_i'\right]^{-1}$$

Income Data



Exponential Regression

```
--> logl     ; lhs=hhninc ; rhs = x ; model=exp $
Normal exit:  11 iterations. Status=0. F=   -1550.075
-----------------------------------------------------------------------
Exponential (Loglinear) Regression Model
Dependent variable                  HHNINC
Log likelihood function       1550.07536
Restricted log likelihood     1195.06953
Chi squared [    5 d.f.]       710.01166
Significance level                .00000
McFadden Pseudo R-squared      -.2970587
Estimation based on N =   27322, K =   6
--------+--------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+--------------------------------------------------------------
        |Parameters in conditional mean function
Constant|    1.77430***       .04501       39.418   .0000
     AGE|     .00205***       .00063        3.274   .0011      43.5272
    EDUC|    -.05572***       .00271      -20.539   .0000      11.3202
 MARRIED|    -.26341***       .01568      -16.804   .0000       .75869
  HHKIDS|     .06512***       .01399        4.657   .0000       .40272
  FEMALE|    -.00542          .01234        -.439   .6603       .47881
--------+--------------------------------------------------------------
Note: ***, **, * = Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------
```

Variance Estimators

```
histogram;rhs=hhninc$
reject ; hhninc=0$
namelist ; X = one, age,educ,married,hhkids,female $
loglinear  ; lhs=hhninc ; rhs = x ; model=exp $
create ; thetai = exp(b'x) $
create ; gi = (hhninc/thetai - 1) ; gi2 = gi^2 $$
create ; hi = (hhninc/thetai) $
matrix ; Expected = <X'X>      ; Stat(b,Expected,X)$
matrix ; Actual   = <X'[hi]X>  ; Stat(b,Actual,X) $
matrix ; BHHH     = <X'[gi2]X> ; Stat(b,BHHH,X) $
matrix ; Robust = Actual * X'[gi2]X * Actual
                               ; Stat(b,Robust,X) $
```

Estimates

```
--------+--------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]
--------+--------------------------------------------------
--> matrix ; Expected = <X'X> ; Stat(b,Expected,X)$
Constant|     1.77430***        .04548       39.010   .0000
     AGE|      .00205***        .00061        3.361   .0008
    EDUC|     -.05572***        .00269      -20.739   .0000
 MARRIED|     -.26341***        .01558      -16.902   .0000
  HHKIDS|      .06512***        .01425        4.571   .0000
  FEMALE|     -.00542           .01235        -.439   .6605
--> matrix ; Actual   = <X'[hi]X> ; Stat(b,Actual,X) $
Constant|     1.77430***        .11922       14.883   .0000
     AGE|      .00205           .00181        1.137   .2553
    EDUC|     -.05572***        .00631       -8.837   .0000
 MARRIED|     -.26341***        .04954       -5.318   .0000
  HHKIDS|      .06512*          .03920        1.661   .0967
  FEMALE|     -.00542           .03471        -.156   .8759
--> matrix ; BHHH     = <X'[gi2]X> ; Stat(b,BHHH,X) $
Constant|     1.77430***        .05409       32.802   .0000
     AGE|      .00205***        .00069        2.973   .0029
    EDUC|     -.05572***        .00331      -16.815   .0000
 MARRIED|     -.26341***        .01737      -15.165   .0000
  HHKIDS|      .06512***        .01637        3.978   .0001
  FEMALE|     -.00542           .01410        -.385   .7004
--> matrix ; Robust = Actual * X'[gi2]X * Actual $
Constant|     1.77430***        .28500        6.226   .0000
     AGE|      .00205           .00481         .427   .6691
    EDUC|     -.05572***        .01306       -4.268   .0000
 MARRIED|     -.26341*          .14581       -1.806   .0708
  HHKIDS|      .06512           .09459         .689   .4911
  FEMALE|     -.00542           .08580        -.063   .9496
```

Testing Hypotheses

Wald tests, using the familiar distance measure

Likelihood ratio tests:

LogL_U = log likelihood without restrictions

LogL_R = log likelihood with restrictions

$\text{LogL}_U > \text{logL}_R$ for any nested restrictions

$2(\text{LogL}_U - \text{logL}_R) \rightarrow$ chi-squared [J]

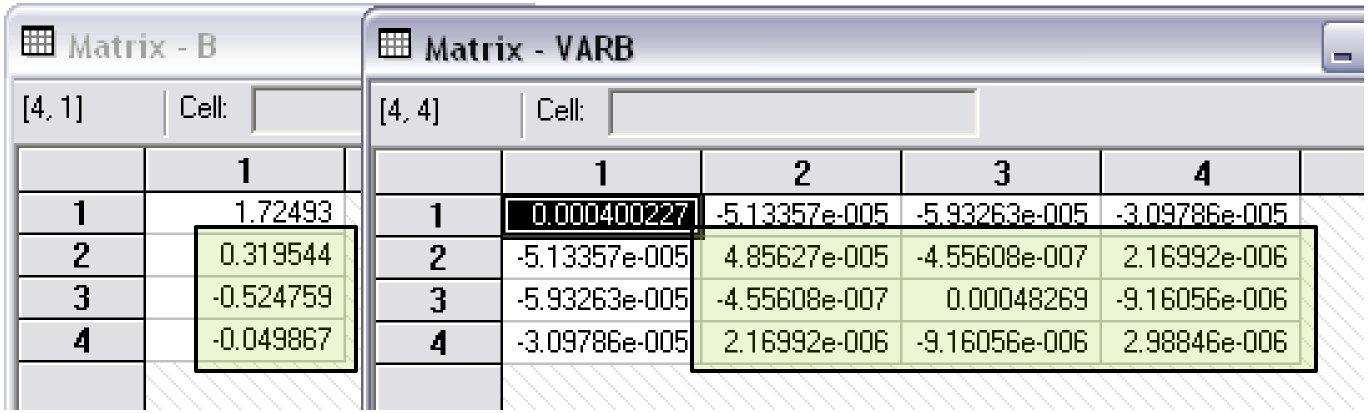
Testing the Model

```
+----------------------------------------------+
| Poisson Regression                            |
| Maximum Likelihood Estimates                  |
| Dependent variable                 DOCVIS     |
| Number of observations              27326     |
| Iterations completed                    7     |
| Log likelihood function       -106215.1       | Log likelihood
| Number of parameters                    4     |
| Restricted log likelihood     -108662.1       | Log Likelihood with only a
| McFadden Pseudo R-squared      .0225193       | constant term.
| Chi squared                    4893.983       | 2*[logL – logL(0)]
| Degrees of freedom                      3     |
| Prob[ChiSqd > value] =         .0000000       |
+----------------------------------------------+
```

Likelihood ratio test that all three slopes are zero.

NOTE: -2logL reported by some computer programs is meaningless.

Wald Test



Matrix - B

[4, 1] Cell:

	1
1	1.72493
2	0.319544
3	-0.524759
4	-0.049867

Matrix - VARB

[4, 4] Cell:

	1	2	3	4
1	0.000400227	-5.13357e-005	-5.93263e-005	-3.09786e-005
2	-5.13357e-005	4.85627e-005	-4.55608e-007	2.16992e-006
3	-5.93263e-005	-4.55608e-007	0.00048269	-9.16056e-006
4	-3.09786e-005	2.16992e-006	-9.16056e-006	2.98846e-006

--> MATRIX ; List ; b1 = b(2:4) ; v11 = varb(2:4,2:4) ; B1'<V11>B1$

```
Matrix B1                     Matrix V11
has  3 rows and  1 columns.   has  3 rows and  3 columns
1                                         1             2             3
+--------------               +---------------------------------------
1|     .31954                1|  .4856275D-04 -.4556076D-06  .2169925D-05
2|    -.52476                2| -.4556076D-06       .00048  -.9160558D-05
3|    -.04987                3|  .2169925D-05 -.9160558D-05  .2988465D-05

Matrix Result   has  1 rows and  1 columns.
              1
       +--------------
      1| 4682.38779       LR statistic was 4893.983
```

Chow Style Test for Structural Change

Does the same model apply to 2 (G) groups?

For linear regression we used the "Chow" (F) test.

For models fit by maximum likelihood, we use a test based on the likelihood function. The same model is fit to the pooled sample and to each group.

$$\text{Chi squared} = 2\left[\left(\sum_{g=1}^{G} \log L_g\right) - \log L_{pooled}\right]$$

Degrees of freedom = (G-1)K.

Poisson Regressions

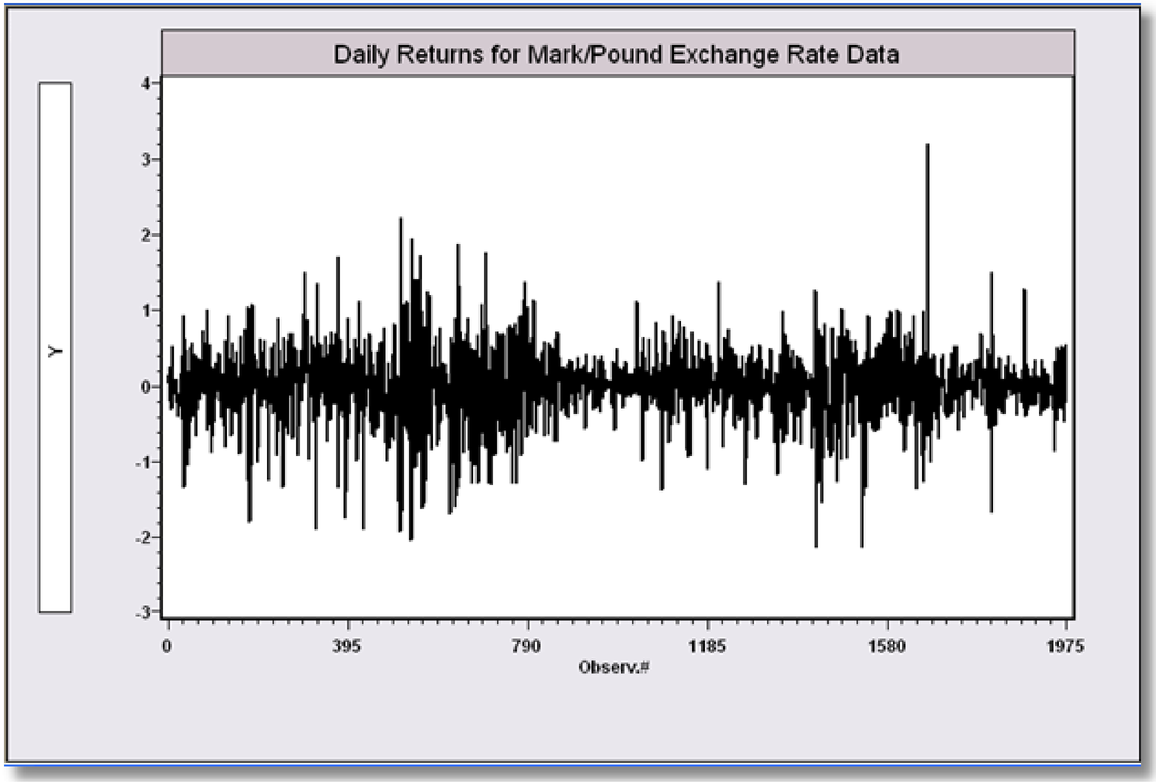
Poisson Regression
Dependent variable DOCVIS
Log likelihood function -90878.20153 (Pooled, N = 27326)
Log likelihood function -43286.40271 (Male, N = 14243)
Log likelihood function -46587.29002 (Female, N = 13083)

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Pooled					
Constant	2.54579***	.02797	91.015	.0000	
AGE	.00791***	.00034	23.306	.0000	43.5257
EDUC	-.02047***	.00170	-12.056	.0000	11.3206
HSAT	-.22780***	.00133	-171.350	.0000	6.78543
HHNINC	-.26255***	.02143	-12.254	.0000	.35208
HHKIDS	-.12304***	.00796	-15.464	.0000	.40273
Males					
Constant	2.38138***	.04053	58.763	.0000	
AGE	.01232***	.00050	24.738	.0000	42.6528
EDUC	-.02962***	.00253	-11.728	.0000	11.7287
HSAT	-.23754***	.00202	-117.337	.0000	6.92436
HHNINC	-.33562***	.03357	-9.998	.0000	.35905
HHKIDS	-.10728***	.01166	-9.204	.0000	.41297
Females					
Constant	2.48647***	.03988	62.344	.0000	
AGE	.00379***	.00048	7.940	.0000	44.4760
EDUC	.00893***	.00234	3.821	.0001	10.8764
HSAT	-.21724***	.00177	-123.029	.0000	6.63417
HHNINC	-.22371***	.02767	-8.084	.0000	.34450
HHKIDS	-.14906***	.01107	-13.463	.0000	.39158

Chi Squared Test

```
Namelist; X = one,age,educ,hsat,hhninc,hhkids$
Sample  ; All $
Poisson ; Lhs = Docvis ; Rhs = X $
Calc    ; Lpool = logl $
Poisson ; For [female = 0] ; Lhs = Docvis ; Rhs = X $
Calc    ; Lmale = logl $
Poisson ; For [female = 1] ; Lhs = Docvis ; Rhs = X $
Calc    ; Lfemale = logl $
Calc    ; K = Col(X) $
Calc    ; List
        ; Chisq = 2*(Lmale + Lfemale - Lpool)
        ; Ctb(.95,k) $
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
CHISQ   =   2009.017601
*Result*=     12.591587
The hypothesis that the same model applies to men and
women is rejected.
```

GARCH Models: A Model for Time Series with Latent Heteroscedasticity



Bollerslev/Ghysel, 1974

ARCH Model

Model Formulation: The mean is stable over time, the variance definitely is not.

A first step. ARCH model.

$y_t = \beta' x_t + \varepsilon_t$

$E[\varepsilon_t] = 0$, $Var[\varepsilon_t] = \sigma^2$ ==> A classical regression model.

ARCH: $E[\varepsilon_t^2 \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots] = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \ldots$ (Autoregressive Conditional Heteroscedasticity)

Using iterated expectations, $E[\varepsilon_t^2] = \alpha_0 + E[\varepsilon_{t-1}^2] + \ldots$

Assume that the unconditional variance in all periods are the same (stationarity). Then,

$\sigma^2 = \alpha_0 / (1 - \alpha_1 - \ldots)$

How many lags?
Researchers found that many were needed accurately to track high frequency (financial) data.

GARCH Model

A better model: GARCH model

GARCH: $E[\varepsilon_t^2 \mid \varepsilon_{t-1}] = \sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \sigma_{t-1}^2$

Turns out to work better and require many fewer parameters

How to estimate:

Distribution: $\varepsilon_t \mid$ past data ~ Normal[$\beta'x_t$, σ_2^2]

Density = $(2\pi)^{-1/2} (\sigma_t^2)^{-1/2} \exp[-(\varepsilon_t/\sigma_t)^2/2]$

Log likelihood: $-T/2 \log(2\pi) - 1/2\Sigma_t \log\sigma_t^2 - 1/2 \Sigma_t \varepsilon_t^2 / \sigma_t^2$

(Note appearance of lagged values. This must be computed recursively from some set of initial values: Current practice uses the sum of squares

$$v_0 = (1/T)\Sigma_t (y_t - b(k)'x_t)^2$$

based on the current values of the parameters, as the initial values for ε_0^2 and σ_0^2

Estimated GARCH Model

```
----------------------------------------------------------------------
GARCH MODEL
Dependent variable                     Y
Log likelihood function     -1106.60788
Restricted log likelihood   -1311.09637
Chi squared [   2 d.f.]       408.97699
Significance level               .00000
McFadden Pseudo R-squared      .1559676
Estimation based on N =   1974, K =   4
GARCH Model, P = 1, Q = 1
Wald statistic for GARCH =     3727.503
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
        |Regression parameters
Constant|    -.00619          .00873        -.709   .4783
        |Unconditional Variance
Alpha(0)|     .01076***       .00312        3.445   .0006
        |Lagged Variance Terms
Delta(1)|     .80597***       .03015       26.731   .0000
        |Lagged Squared Disturbance Terms
Alpha(1)|     .15313***       .02732        5.605   .0000
        |Equilibrium variance, a0/[1-D(1)-A(1)]
EquilVar|     .26316          .59402         .443   .6577
--------+-------------------------------------------------------------
```

2 Step Estimation (Murphy-Topel)

Setting, fitting a model which contains parameter estimates from another model.

Typical application, inserting a prediction from one model into another.

A. Procedures: How it's done.

B. Asymptotic results:

   1. Consistency
   2. Getting an appropriate estimator of the asymptotic covariance matrix

The Murphy - Topel result

Application: Equation 1: Number of children

Equation 2: Labor force participation

Setting

- Two equation model:
  - Model for $y_1 = f(y_1 \mid \mathbf{x}_1, \boldsymbol{\theta}_1)$
  - Model for $y_2 = f(y_2 \mid \mathbf{x}_2, \boldsymbol{\theta}_2, \mathbf{x}_1, \boldsymbol{\theta}_1)$
  - (Note, not ‘simultaneous’ or even ‘recursive.’)
- Procedure:
  - Estimate $\boldsymbol{\theta}_1$ by ML, with covariance matrix $(1/n)\mathbf{V}_1$
  - Estimate $\boldsymbol{\theta}_2$ by ML treating $\boldsymbol{\theta}_1$ as if it were known.
  - Correct the estimated asymptotic covariance matrix, $(1/n)\mathbf{V}_2$ for the estimator of $\boldsymbol{\theta}_2$

Murphy and Topel (1984,2002) Results

Both MLEs are consistent

$$\text{Asy.Var}[\hat{\theta}_2] = \tfrac{1}{n}\left[\mathbf{V}_2 + \mathbf{V}_2(\mathbf{C}\mathbf{V}_1\mathbf{C}' - \mathbf{R}\mathbf{V}_1\mathbf{C}' - \mathbf{C}\mathbf{V}_1\mathbf{R}')\mathbf{V}_2\right]$$

$$\mathbf{V}_1 = \text{Asy.Var}\sqrt{n}[\hat{\theta}_1 - \theta_1]$$

$$\mathbf{V}_2 = \text{Asy.Var}\sqrt{n}[\hat{\theta}_2 - \theta_2] \mid \theta_1$$

$$\mathbf{C} = E\left[\frac{1}{n}\left(\frac{\partial \log L_2}{\partial \theta_2}\right)\left(\frac{\partial \log L_2}{\partial \theta_1'}\right)\right]$$

$$\mathbf{R} = E\left[\frac{1}{n}\left(\frac{\partial \log L_2}{\partial \theta_2}\right)\left(\frac{\partial \log L_1}{\partial \theta_1'}\right)\right]$$

M&T Computations

First equation: $\hat{\theta}_1$=MLE,

$$\hat{\mathbf{V}}_1 = \left[\tfrac{1}{n}\Sigma_{i=1}^{N} -\hat{\mathbf{H}}_{i1}\right]^{-1} \quad \text{or} \quad \left[\tfrac{1}{n}\Sigma_{i=1}^{N}\hat{\mathbf{g}}_{i1}\hat{\mathbf{g}}_{i1}'\right]^{-1}$$

Second equation: $\hat{\theta}_2$=MLE| $\hat{\theta}_1$

$$\hat{\mathbf{V}}_2 = \left[\tfrac{1}{n}\Sigma_{i=1}^{N} -\hat{\mathbf{H}}_{i2}\right]^{-1} \quad \text{or} \quad \left[\tfrac{1}{n}\Sigma_{i=1}^{N}\hat{\mathbf{g}}_{i2}\hat{\mathbf{g}}_{i3}'\right]^{-1}$$

$$\mathbf{C} = \frac{1}{n}\Sigma_{i=1}^{N}\left(\frac{\partial \ln f_2(y_2 \mid \mathbf{x}_2,\hat{\theta}_2,\mathbf{x}_1,\hat{\theta}_1)}{\partial\hat{\theta}_2}\right)\left(\frac{\partial \ln f_2(y_2 \mid \mathbf{x}_2,\hat{\theta}_2,\mathbf{x}_1,\hat{\theta}_1)}{\partial\hat{\theta}_1}\right)$$

$$\mathbf{R} = \frac{1}{n}\Sigma_{i=1}^{N}\left(\frac{\partial \ln f_2(y_2 \mid \mathbf{x}_2,\hat{\theta}_2,\mathbf{x}_1,\hat{\theta}_1)}{\partial\hat{\theta}_2}\right)\left(\frac{\partial \ln f_1(y_1 \mid \mathbf{x}_1,\hat{\theta}_1)}{\partial\hat{\theta}_1}\right)$$

Example

Equation 1: Number of Kids – Poisson Regression

- $p(y_{i1}|\mathbf{x}_{i1}, \boldsymbol{\beta})=\exp(-\lambda_i)\lambda_i^{y_{i1}}/y_{i1}!$
- $\lambda_i = \exp(\mathbf{x}_{i1}'\boldsymbol{\beta})$
- $\mathbf{g}_{i1} = \mathbf{x}_{i1}(y_{i1} - \lambda_i)$
- $\mathbf{V}_1 = [(1/n)\Sigma(-\lambda_i)\mathbf{x}_{i1}\mathbf{x}_{i1}']^{-1}$

Example - Continued

Equation 2: Labor Force Participation – Logit

$$p(y_{i2}|\mathbf{x}_{i2},\boldsymbol{\delta},\alpha,\mathbf{x}_{i1},\boldsymbol{\beta})=\exp(d_{i2})/[1+\exp(d_{i2})]=P_{i2}$$

$$d_{i2} = (2y_{i2}-1)[\boldsymbol{\delta}'\mathbf{x}_{i2} + \alpha\lambda_i]$$

$$\lambda_i \ = \exp(\boldsymbol{\beta}'\mathbf{x}_{i1})$$

Let $\mathbf{z}_{i2} = (\mathbf{x}_{i2}, \lambda_i)$, $\boldsymbol{\theta}_2 = (\boldsymbol{\delta}, \alpha)$

$$d_{i2} \ = (2y_{i2}-1)[\boldsymbol{\theta}_2'\mathbf{z}_{i2}]$$

$$\mathbf{g}_{i2} \ = (y_{i2}-P_{i2})\mathbf{z}_{i2}$$

$$\mathbf{V}_2 \ = [(1/n)\Sigma\{-P_{i2}(1-P_{i2})\}\mathbf{z}_{i2}\mathbf{z}_{i2}']^{-1}$$

Murphy and Topel Correction

$$\mathbf{C} = \frac{1}{N}\Sigma_{i=1}^{N}\left[(y_{i2} - P_{i2})\mathbf{z}_{i2}\right]\left[(y_{i2} - P_{i2})\alpha\lambda_i\mathbf{x}_{i1}\right]$$

$$\mathbf{R} = \frac{1}{N}\Sigma_{i=1}^{N}\left[(y_{i2} - P_{i2})\mathbf{z}_{i2}\right]\left[(y_{i1} - \lambda_i)\mathbf{x}_{i1}\right]$$

Two Step Estimation of LFP Model

```
? Data transformations.  Number of kids, scale income variables
Create ; Kids = kl6 + k618
       ; income = faminc/10000 ; Wifeinc = ww*whrs/1000 $
? Equation 1, number of kids.  Standard Poisson fertility model.
? Fit equation, collect parameters BETA and covariance matrix V1
? Then compute fitted values and derivatives
Namelist ; X1  = one,wa,we,income,wifeinc$
Poisson  ; Lhs = kids  ; Rhs = X1 $
Matrix   ; Beta = b  ; V1 = N*VARB $
Create   ; Lambda = Exp(X1'Beta); gi1  = Kids - Lambda $
? Set up logit labor force participation model
? Compute probit model and collect results.  Delta=Coefficients on X2
? Alpha = coefficient on fitted number of kids.
Namelist ; X2 = one,wa,we,ha,he,income  ;  Z2 = X2,Lambda $
Logit    ; Lhs = lfp   ; Rhs = Z2 $
Calc     ; alpha = b(kreg) ; K2 = Col(X2) $
Matrix   ; delta=b(1:K2)   ; Theta2 = b  ; V2 = N*VARB $
? Poisson derivative of with respect to beta is (kidsi - lambda)´X1
Create   ; di = delta'X2 + alpha*Lambda
         ; pi2= exp(di)/(1+exp(di))
         ; gi2  = LFP - Pi2
? These are the terms that are used to compute R and C.
         ; ci = gi2*gi2*alpha*lambda
         ; ri = gi2*gi1$
MATRIX   ; C = 1/n*Z2'[ci]X1
         ; R = 1/n*Z2'[ri]X1
         ; A = C*V1*C' - R*V1*C' - C*V1*R'
         ; V2S = V2+V2*A*V2 ; V2s = 1/N*V2S $
? Compute matrix products and report results
Matrix  ; Stat(Theta2,V2s,Z2)$
```

Estimated Equation 1: E[Kids]

```
+----------------------------------------------+
| Poisson Regression                           |
| Dependent variable                 KIDS      |
| Number of observations              753      |
| Log likelihood function       -1123.627      |
+----------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant      3.34216852       .24375192     13.711    .0000
 WA            -.06334700       .00401543    -15.776    .0000   42.5378486
 WE            -.02572915       .01449538     -1.775    .0759   12.2868526
 INCOME         .06024922       .02432043      2.477    .0132   2.30805950
 WIFEINC       -.04922310       .00856067     -5.750    .0000   2.95163126
```

Two Step Estimator

Multinomial Logit Model	
Dependent variable	LFP
Number of observations	753
Log likelihood function	-351.5765
Number of parameters	7

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Characteristics in numerator of Prob[Y = 1]					
Constant	33.1506089	2.88435238	11.493	.0000	
WA	-.54875880	.05079250	-10.804	.0000	42.5378486
WE	-.02856207	.05754362	-.496	.6196	12.2868526
HA	-.01197824	.02528962	-.474	.6358	45.1208499
HE	-.02290480	.04210979	-.544	.5865	12.4913679
INCOME	.39093149	.09669418	4.043	.0001	2.30805950
LAMBDA	-5.63267225	.46165315	-12.201	.0000	1.59096946

With Corrected Covariance Matrix

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]
Constant	33.1506089	5.41964589	6.117	.0000
WA	-.54875880	.07780642	-7.053	.0000
WE	-.02856207	.12508144	-.228	.8194
HA	-.01197824	.02549883	-.470	.6385
HE	-.02290480	.04862978	-.471	.6376
INCOME	.39093149	.27444304	1.424	.1543
LAMBDA	-5.63267225	1.07381248	-5.245	.0000

Optimization - Algorithms

Maximize or minimize (optimize) a function F(θ)

Algorithm = rule for searching for the optimizer

Iterative algorithm: $\theta^{(k+1)} = \theta^{(k)} + \text{Update}^{(k)}$

Derivative (gradient) based algorithm

$$\theta^{(k+1)} = \theta^{(k)} + \text{Update}(\mathbf{g}^{(k)})$$

Update is a function of the gradient.

Compare to 'derivative free' methods

(Discontinuous criterion functions)

Newton's Method

Computing the Maximum Likelihood Estimator - An Iteration

We used $\left(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}\right) \approx [-\mathbf{H}(\boldsymbol{\theta})]^{-1}\,\mathbf{g}(\boldsymbol{\theta})$ to derive the asymptotic properties.

Consider using this as a method of finding the actual maximum of the likelihood function. Suppose $\boldsymbol{\theta}_0$ is an ***estimate***, of the true parameter, not the actual one. Then, it is also true that

$$\hat{\boldsymbol{\theta}}_{ML} \approx \boldsymbol{\theta}_0 - [\mathbf{H}(\boldsymbol{\theta}_0)]^{-1}\,\mathbf{g}(\boldsymbol{\theta}_0)$$

That is, if we have this estimate $\boldsymbol{\theta}_0$, approximate the MLE by using this update formula, just by computing the derivatives at $\boldsymbol{\theta}_0$. (Puzzle enthusiasts: Show that this scheme works perfectly with one iteration for the linear regression model for any vector $\boldsymbol{\theta}_0$ you choose.) This iteration is called Newton's method. In fact, it is a general method of finding the extremum of a function. Once we "update" $\boldsymbol{\theta}_0$, redo the computation, using the updated vector as the next $\boldsymbol{\theta}_0$.

Newton's Method for Poisson Regression

The likelihood equation is

$$\Sigma_i \, [\mathbf{x}_i \, (y_i - \exp(\beta' x_i))] = 0$$

How to solve this nonlinear set of equations to compute the maximum likelihood estimator? Use Newton's method. Recall how we used this to do nonlinear least squares

$$b(k+1) = b(k) + [\text{Hessian}(k)]^{-1} \, [\text{derivative}(k)]$$

where (k) means evaluate at b(k), kth iteration

$$b(k+1) = b(k) - [X'\Lambda X]^{-1} X'e$$

This problem is 'globally concave.' The Hessian is negative definite. We can use any starting value. 0 usually works. There is only one solution.

Poisson Regression Iterations

```
Poisson ; lhs = doctor ; rhs = one,female,hhninc,educ;mar;output=3$
Method=Newton; Maximum iterations=100
Convergence criteria: gtHg   .1000D-05 chg.F   .0000D+00 max|db|   .0000D+00
Start values:    .00000D+00   .00000D+00   .00000D+00   .00000D+00
1st derivs.    -.13214D+06  -.61899D+05  -.43338D+05  -.14596D+07
Parameters:     .28002D+01   .72374D-01  -.65451D+00  -.47608D-01
Itr  2 F= -.1587D+06 gtHg=  .2832D+03 chg.F=  .1587D+06 max|db|=  .1346D+01
1st derivs.    -.33055D+05  -.14401D+05  -.10804D+05  -.36592D+06
Parameters:     .21404D+01   .16980D+00  -.60181D+00  -.48527D-01
Itr  3 F= -.1115D+06 gtHg=  .9725D+02 chg.F=  .4716D+05 max|db|=  .6348D+00
1st derivs.    -.42953D+04  -.15074D+04  -.13927D+04  -.47823D+05
Parameters:     .17997D+01   .27758D+00  -.54519D+00  -.49513D-01
Itr  4 F= -.1063D+06 gtHg=  .1545D+02 chg.F=  .5162D+04 max|db|=  .1437D+00
1st derivs.    -.11692D+03  -.22248D+02  -.37525D+02  -.13159D+04
Parameters:     .17276D+01   .31746D+00  -.52565D+00  -.49852D-01
Itr  5 F= -.1062D+06 gtHg=  .5006D+00 chg.F=  .1218D+03 max|db|=  .6542D-02
1st derivs.    -.12522D+00  -.54690D-02  -.40254D-01  -.14232D+01
Parameters:     .17249D+01   .31954D+00  -.52476D+00  -.49867D-01
Itr  6 F= -.1062D+06 gtHg=  .6215D-03 chg.F=  .1254D+00 max|db|=  .9678D-05
1st derivs.    -.19317D-06  -.94936D-09  -.62872D-07  -.22029D-05
Parameters:     .17249D+01   .31954D+00  -.52476D+00  -.49867D-01
Itr  7 F= -.1062D+06 gtHg=  .9957D-09 chg.F=  .1941D-06 max|db|=  .1602D-10
                        * Converged
```

Optimization

Algorithms

Iteration $\theta^{(k+1)} = \theta^{(k)} + \text{Update}^{(k)}$

General structure: $\theta^{(k+1)} = \theta^{(k)} + \lambda^{(k)}\mathbf{W}^{(k)}\mathbf{g}^{(k)}$

$\mathbf{g}^{(k)}$ = derivative vector, points to a better value than $\theta^{(k)}$

= direction vector

$\lambda^{(k)}$ = 'step size'

$\mathbf{W}^{(k)}$= a weighting matrix

Algorithms are defined by the choices of $\lambda^{(k)}$ and $\mathbf{W}^{(k)}$

Algorithms

Steepest Ascent: $\lambda^{(k)} = \dfrac{-\mathbf{g}^{(k)\prime}\mathbf{g}^{(k)}}{\mathbf{g}^{(k)\prime}\mathbf{H}^{(k)}\mathbf{g}^{(k)}}$, $\mathbf{W}^{(k)} = \mathbf{I}$

$\mathbf{g}^{(k)}$ = first derivative vector

$\mathbf{H}^{(k)}$ = second derivatives matrix

Newton's Method: $\lambda^{(k)} = -1$, $\mathbf{W}^{(k)} = [\mathbf{H}^{(k)}]^{-1}$

(Sometimes called Newton-Raphson.

Method of Scoring: $\lambda^{(k)} = -1$, $\mathbf{W}^{(k)} = [E[\mathbf{H}^{(k)}]]^{-1}$

(Scoring uses the expected Hessian. Usually inferior to Newton's method. Takes more iterations.)

BHHH Method (for MLE): $\lambda^{(k)} = -1$, $\mathbf{W}^{(k)} = [\Sigma_{i=1}^{n}\mathbf{g}_i^{(k)}\mathbf{g}_i^{(k)\prime}]^{-1}$

Line Search Methods

Squeezing: Essentially trial and error

$\lambda^{(k)} =$ 1, 1/2, 1/4, 1/8, ...

Until the function improves

Golden Section: Interpolate between $\lambda^{(k)}$ and $\lambda^{(k-1)}$

Others : Many different methods have been suggested

Quasi-Newton Methods

How to construct the weighting matrix:

Variable metric methods:

$$\mathbf{W}^{(k)} = \mathbf{W}^{(k-1)} + \mathbf{E}^{(k-1)}, \ \mathbf{W}^{(1)} = \mathbf{I}$$

Rank one updates: $\mathbf{W}^{(k)} = \mathbf{W}^{(k-1)} + \mathbf{a}^{(k-1)}\mathbf{a}^{(k-1)\prime}$

(Davidon Fletcher Powell)

There are rank two updates (Broyden) and higher.

Stopping Rule

When to stop iterating: 'Convergence'

(1) Derivatives are small? Not good.

Maximizer of F() is the same as that of .0000001F(), but the derivatives are small right away.

(2) Small absolute change in parameters from one iteration to the next? Problematic because it is a function of the stepsize which may be small.

(3) Commonly accepted 'scale free' measure

$$\Delta = g^{(k)\prime}[H^{(k)}]^{-1}g^{(k)}$$

```
-----------------------------------------------------------
Nonlinear Estimation of Model Parameters
Method                                         BFGS
Maximum iterations                              100
Maximum steps in line search                     20
Convergence criteria
  Gradient norm                          .10000D-05
  Change in function value               .00000D+00
  Maximum parameter change               .00000D+00
Laguerre                                         20
Hermite                                          64
Replications for GHK simulator=                 100
Starting Values-------------------------------------
              B(001)          .104374021D+01
              B(002)          .000000000D+00
              B(003)          .000000000D+00
              B(004)          .000000000D+00
              B(005)          .100000000D+01
===========================================================
Current iteration    1
Parameter                  Value        Derivative
B(001)          .104374D+01         .339114D-09
B(002)          .000000D+00         .838517D+04
B(003)          .000000D+00         .983928D+03
B(004)          .000000D+00        -.919225D+03
B(005)          .100000D+01        -.126660D+05
Function value                         -.119525D+04
Convergence criteria at iteration   1...
Gradient norm                           .152496D+05
Function change                         .119525D+04
Maximum parameter change                .838517D+10
```

Start value for constant is log (mean HHNINC)

Start value for P is 1 => exponential model.

```
===========================================================
Current iteration    1
Parameter            Value        Derivative
B(001)         .104374D+01      .339114D-09
B(002)         .000000D+00      .838517D+04
B(003)         .000000D+00      .983928D+03
B(004)         .000000D+00     -.919225D+03
B(005)         .100000D+01     -.126660D+05
Function value                 -.119525D+04
Convergence criteria at iteration   1...
Gradient norm                   .152496D+05
Function change                 .119525D+04
Maximum parameter change        .838517D+10
Line Search------------------------------------------------
Try =  0 F= -.1195D+04 Step=  .0000D+00 Slope= -.1525D+05
Try =  1 F=  .2519D+03 Step=  .1000D+00 Slope=  .4106D+05
Try =  2 F= -.1375D+04 Step=  .2404D-01 Slope=  .8372D+02
===========================================================
Current iteration    2
Parameter            Value        Derivative
B(001)         .104374D+01     -.292653D+04
B(002)        -.132171D-01     -.276484D+05
B(003)        -.155092D-02     -.126624D+04
B(004)         .144893D-02     -.122874D+06
B(005)         .101996D+01     -.938394D+04
Function value                 -.137485D+04
Convergence criteria at iteration   2...
Gradient norm                   .126336D+06
Function change                 .179602D+03
Maximum parameter change        .848034D+08
Line Search------------------------------------------------
Try =  0 F= -.1375D+04 Step=  .0000D+00 Slope= -.1263D+06
Try =  1 F=  .1963D+05 Step=  .2404D-01 Slope=  .2372D+07
Try =  2 F= -.1515D+04 Step=  .2604D-02 Slope=  .2167D+05
Try =  3 F= -.1519D+04 Step=  .2243D-02 Slope= -.1881D+02
===========================================================
```

```
===========================================================
Current iteration   18
Parameter           Value       Derivative
B(001)         .345578D+01    -.889814D-02
B(002)        -.554363D-01    -.573045D-01
B(003)        -.236661D+00    -.156585D-01
B(004)         .874294D-03    -.461608D+00
B(005)         .510591D+01     .212071D-03
Function value                -.142407D+05
Convergence criteria at iteration  18...
Gradient norm                  .101119D-03
Function change                .292575D-04
Maximum parameter change       .141822D-05
Line Search-------------------------------------------------
Try =  0 F= -.1424D+05 Step=  .0000D+00 Slope= -.3444D-02
Try =  1 F= -.1424D+05 Step=  .1486D-03 Slope=  .1715D+00
Try =  2 F= -.1424D+05 Step=  .2926D-05 Slope= -.6930D-06
===========================================================
Current iteration   19
Parameter           Value       Derivative
B(001)         .345579D+01    -.405655D-04
B(002)        -.554364D-01    -.618229D-03
B(003)        -.236660D+00    -.636325D-05
B(004)         .874293D-03    -.720424D-03
B(005)         .510591D+01     .209628D-05
Function value                -.142407D+05
Convergence criteria at iteration  19...
Gradient norm                  .411634D-06[Grdnt Converged]
Function change                .503132D-08
Maximum parameter change       .594844D-07
Iterative procedure has converged
-----------------------------------------------------------
Normal exit:  19 iterations. Status=0, F=   -.1424074D+05
Function value was  -.1195245094D+04 at entry -----------
                    -.1424074142D+05 at exit  -----------
```

```
----------------------------------------------------------------------------
Gamma (Loglinear) Regression Model
Dependent variable                HHNINC
Log likelihood function      14240.74142
Restricted log likelihood     1195.24508
Chi squared [   4](P= .000)  26090.99269
Significance level                .00000
McFadden Pseudo R-squared    -10.9144949
Estimation based on N =   27326, K =   5
Inf.Cr.AIC  = -28471.5 AIC/N =    -1.042
--------+-------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  HHNINC|  Coefficient     Error        z    |z|>Z*        Interval
--------+-------------------------------------------------------------------
        |Parameters in conditional mean function...........................
Constant|    -3.45579***      .02043  -169.15  .0000    -3.49583  -3.41574
    EDUC|      .05544***      .00118    46.89  .0000      .05312    .05775
 MARRIED|      .23666***      .00646    36.61  .0000      .22399    .24933
     AGE|     -.00087***      .00025    -3.45  .0006     -.00137   -.00038
        |Scale parameter for gamma model...................................
 P_scale|     5.10591***      .04233   120.62  .0000     5.02294   5.18888
--------+-------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
Model was estimated on Mar 22, 2017 at 04:44:24 PM
----------------------------------------------------------------------------
```