Advances in Count Data Regression:
I. Basic cross-section methods

A. Colin Cameron
Univ. of Calif. - Davis

28th Annual Workshop in Applied Statistics
Southern California Chapter of the American Statistical Association
Held at University of California - Los Angeles
http://cameron.econ.ucdavis.edu/racd/count.html

March 28 2009

Introduction

- Count data models are for dependent variable $y = 0, 1, 2, ...$
- Example:
  - y: Number of doctor visits (usually cross-section)
    $\mathbf{x}$: health status, age, gender, ...
- Many approaches and issues are general nonlinear model issues.
  - Econometrics: MLE, quasi-MLE, generalized MM (GMM)
  - Statistics: generalized linear models (GLM).

Outline of all lectures

I. Basic cross-section methods:
   - Poisson, GLM, negative binomial

II. More advanced cross-section methods:
   - Hurdle, zero-inflated, finite mixtures, endogeneity

III. Time series and panel methods

IV. Further Topics:
   - multivariate, maximum simulated likelihood, Bayesian

Key sources

- A. Colin Cameron and Pravin K. Trivedi (2010)
  Regression Analysis of Count Data (RACD)
  Second edition, Cambridge Univ. Press, in preparation.
- A. Colin Cameron and Pravin K. Trivedi (2009)
  Microeconometrics using Stata (MUS), chapter 17, Stata Press.
- A. Colin Cameron and Pravin K. Trivedi (2005)
  Microeconometrics: Methods and Applications (MMA), Cambridge Univ. Press.
- Pravin K. Trivedi and Murat Munkin (2009)
  "Recent Developments in Cross Section and Panel Count Models"
  in D. Giles and A. Ullah eds., *Handbook of Empirical Economics and Finance*, Chapman and Hall / CRC, forthcoming.

Outline of basic cross-section count methods

- Introduction
- Count data without regression
  - Doctor visits data
  - Poisson distribution
  - Negative binomial distribution
- Poisson regression
  - Poisson MLE
  - Coefficient interpretation
  - Marginal effects
- Generalized linear models
- Nonlinear least squares
- Diagnostics
- Negative binomial model

Count data without regression: doctor visits

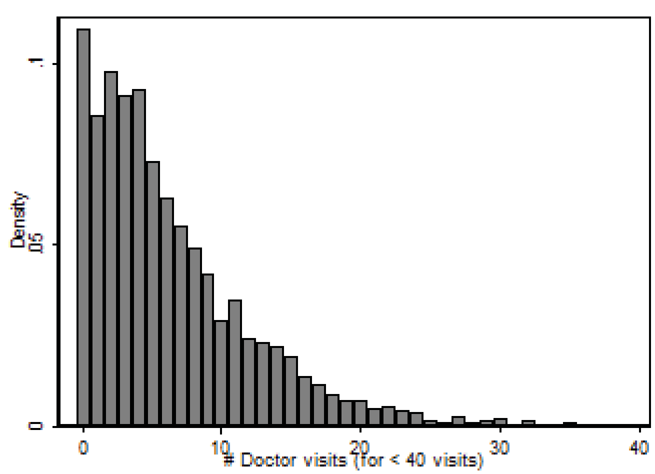
- Many health surveys measure health use as counts as people have better recall of counts than of dollars spent.
- 2003 U.S. Medical Expenditure Panel Survey (MEPS).
- Sample of Medicare population aged 65 and higher (N = 3,677)
- docvis = annual number of doctor visits

```
. use mus17data.dta

. summarize docvis
```

Variable	Obs	Mean	Std. Dev.	Min	Max
docvis	3677	6.822682	7.394937	0	144

Doctor visits: Histogram dropping observations with more than 40 visits



Doctor visits: Frequencies with 11-40 and 41-60 grouped

```
. tabulate dvrange
```

dvrange	Freq.	Percent	Cum.
0	401	10.91	10.91
1	314	8.54	19.45
2	358	9.74	29.18
3	334	9.08	38.26
4	339	9.22	47.48
5	266	7.23	54.72
6	231	6.28	61.00
7	202	5.49	66.49
8	179	4.87	71.36
9	154	4.19	75.55
10	108	2.94	78.49
11-40	774	21.05	99.54
41-60	14	0.38	99.92
73	1	0.03	99.95
106	1	0.03	99.97
144	1	0.03	100.00
Total	3,677	100.00	

Poisson distribution

- From stochastic process theory, natural model for counts is

$$y \sim \textit{Poisson}[\lambda].$$

- Probability mass function:

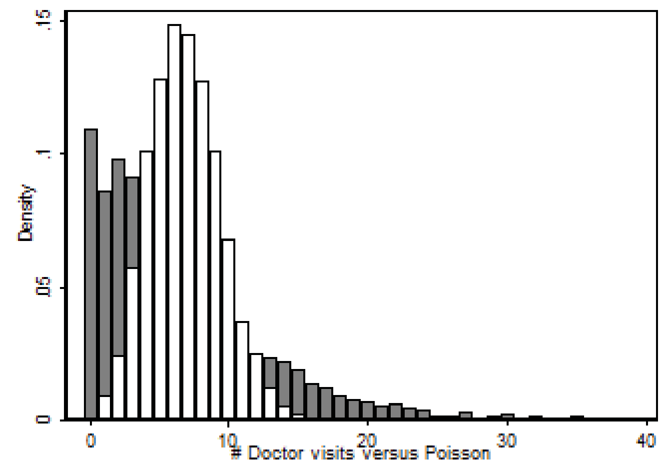
$$\Pr[Y = y|\lambda] = \frac{e^{-\lambda}\lambda^y}{y!}$$

- Mean and variance:

$$\begin{aligned} \mathrm{E}[y] &= \lambda \\ \mathrm{V}[y] &= \lambda \end{aligned}$$

- Equidispersion: variance = mean
  - Restriction imposed by Poisson
- Overdispersion: variance > mean
  - More common feature of count data.

Poisson distribution for $\lambda = \bar{y}$ compared to actual data



Poisson clearly inappropriate: $\bar{y} = 6.82$, $s_y = 7.39$, $s_y^2 = 54.68 \simeq 8.01\bar{y}$.

Negative binomial distribution

- If

$$\begin{aligned} y &\sim \textit{Poisson}[\lambda \upsilon] \\ \upsilon &\sim \textit{Gamma}[\mu = 1, \sigma^2 = \alpha] \end{aligned}$$

then

$$y \sim \textit{Negative Binomial}[\mu = \lambda,\ \sigma^2 = \lambda + \alpha\lambda^2].$$

- Probability mass function:

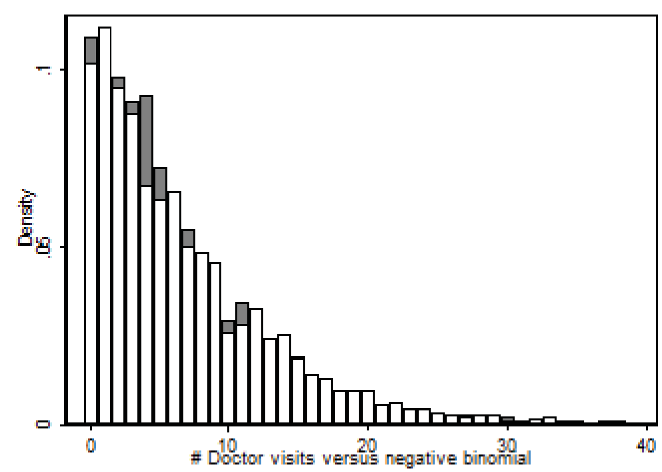
$$\Pr[Y = y|\lambda, \alpha] = \frac{\Gamma(\alpha^{-1} + y)}{\Gamma(\alpha^{-1})\Gamma(y+1)} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda}\right)^{\alpha^{-1}} \left(\frac{\lambda}{\lambda + \alpha^{-1}}\right)^{y}.$$

- Mean and variance:

$$\begin{aligned} \mathrm{E}[y] &= \lambda \\ \mathrm{V}[y] &= \alpha\lambda^2 \end{aligned}$$

- Overdispersion: variance > mean.

Negative binomial for $\lambda = \bar{y}$ and $\alpha = 0.8408$ compared to actual data



Negative binomial much more appropriate than Poisson for these data.

2. Poisson regression: summary

- Poisson regression is straightforward
  - many packages do Poisson regression
  - coefficients are easily interpreted as semi-elasticities.
- Do Poisson rather than OLS with dependent variable
  - y
  - $\ln y$ (with adjustment for $\ln 0$)
  - $\sqrt{y}$ (a variance-stabilizing transformation).
- Poisson MLE is consistent provided only that $\mathrm{E}[y|\mathbf{x}] = \exp(\mathbf{x}'\boldsymbol{\beta})$.
  - But make sure standard errors etc. are robust to $\mathrm{V}[y|\mathbf{x}] \neq \mathrm{E}[y|\mathbf{x}]$.
  - And generally don't use Poisson if need to predict probabilities.

Poisson regression: Poisson MLE

- Let the Poisson rate parameter vary across individuals with $\mathbf{x}$ in way to ensure $\lambda > 0$.

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

- MLE is straightforward given data independent over i.

$$\begin{array}{rll} & f(y) & = e^{-\lambda}\lambda^y/y! \\ \Rightarrow & \ln f(y) & = -\exp(\mathbf{x}'\boldsymbol{\beta}) + y\mathbf{x}'\boldsymbol{\beta} - \ln y! \\ \Rightarrow & \ln L(\boldsymbol{\beta}) & = \sum_{i=1}^n \{-\exp(\mathbf{x}_i'\boldsymbol{\beta}) + y_i\mathbf{x}_i'\boldsymbol{\beta} - \ln y_i!\} \\ \Rightarrow & \frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} & = \sum_{i=1}^n \{-\exp(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i + y_i\mathbf{x}_i\} \end{array}$$

- The ML first-order conditions are

$$\sum_{i=1}^n (y_i - \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}))\mathbf{x}_i = \mathbf{0}.$$

- No explicit solution for $\widehat{\boldsymbol{\beta}}$.
  - Instead use Newton-Raphson iterative method.
  - Fast as objective function is globally concave in $\boldsymbol{\beta}$.

Poisson regression: consistency of Poisson MLE

- ML first-order conditions are

$$\sum_{i=1}^n (y_i - \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}))\mathbf{x}_i = \mathbf{0}.$$

- Consistency only requires (given independence over i)

$$\mathrm{E}[(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))\mathbf{x}_i] = \mathbf{0}$$

- So consistent if

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

- Poisson MLE is consistent if the conditional mean is correctly specified
  - like MLE for linear model under normality.

Poisson regression: distribution of Poisson MLE

- If distribution is Poisson then $\widehat{\boldsymbol{\beta}} \stackrel{a}{\sim} \mathcal{N}[\boldsymbol{\beta}, \mathrm{V}_{\mathrm{MLE}}[\widehat{\boldsymbol{\beta}}]]$ where

$$\widehat{\mathrm{V}}_{\mathrm{MLE}}[\widehat{\boldsymbol{\beta}}] = \left(\sum_i \widehat{\mu}_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1}$$

- If distribution is not Poisson but $\mathrm{E}[y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta})$ and $\mathrm{V}[y_i|\mathbf{x}_i] = \sigma_i^2$ then $\widehat{\boldsymbol{\beta}} \stackrel{a}{\sim} \mathcal{N}[\boldsymbol{\beta}, \mathrm{V}_{\mathrm{ROB}}[\widehat{\boldsymbol{\beta}}]]$ and we use the robust sandwich estimate of variance (White (1982), Huber (1967))

$$\widehat{\mathrm{V}}_{\mathrm{ROB}}[\widehat{\boldsymbol{\beta}}] = \left(\sum_i \widehat{\mu}_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1} \left(\sum_i (y_i - \widehat{\mu}_i)^2 \mathbf{x}_i\mathbf{x}_i'\right) \left(\sum_i \widehat{\mu}_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1}$$

  - $\mathrm{V}_{\mathrm{ROB}}[\widehat{\boldsymbol{\beta}}] = \mathrm{V}_{\mathrm{MLE}}[\widehat{\boldsymbol{\beta}}]$ if $\sigma_i^2 = \mu_i$ (imposed by Poisson)
  - $\mathrm{V}_{\mathrm{ROB}}[\widehat{\boldsymbol{\beta}}] = \alpha \mathrm{V}_{\mathrm{MLE}}[\widehat{\boldsymbol{\beta}}]$ if $\sigma_i^2 = \alpha\mu_i$ (used in GLM literature)
  - Robust se's are much larger than default ML se's if $\alpha >> 1$.

Poisson regression: derivation of robust sandwich

- Take a first-order Taylor series expansion of $\sum_i (y_i - \exp(\mathbf{x}_i'\widehat{\beta}))\mathbf{x}_i$ about β.

$$\sum_i (y_i - \exp(\mathbf{x}_i'\widehat{\beta}))\mathbf{x}_i = \sum_i (y_i - \exp(\mathbf{x}_i'\beta))\mathbf{x}_i - \sum_i \exp(\mathbf{x}_i'\beta)\mathbf{x}_i\mathbf{x}_i'(\widehat{\beta} - \beta)$$

- F.o.c. set this to zero and can show that R disappears asymptotically

$$\begin{aligned} & \textstyle\sum_i (y_i - \mu_i)\mathbf{x}_i + \left(\sum_i -\mu_i \mathbf{x}_i\mathbf{x}_i'\right)(\widehat{\beta} - \beta) = \mathbf{0} \\ \Rightarrow \quad & (\widehat{\beta} - \beta) = \left(\textstyle\sum_i \mu_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1} \times \textstyle\sum_i (y_i - \mu_i)\mathbf{x}_i \\ & \overset{a}{\sim} \left(\textstyle\sum_i \mu_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1} \times \mathcal{N}\left[0, \textstyle\sum_i \sigma_i^2 \mathbf{x}_i\mathbf{x}_i'\right] \\ & \overset{a}{\sim} \mathcal{N}\left[0, \left(\textstyle\sum_i \mu_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1} \left(\textstyle\sum_i \sigma_i^2 \mathbf{x}_i\mathbf{x}_i'\right) \left(\textstyle\sum_i \mu_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1}\right] \end{aligned}$$

where $\mu_i = \exp(\mathbf{x}_i'\beta)$, $\widehat{\mu}_i = \exp(\mathbf{x}_i'\widehat{\beta})$ and $\sigma_i^2 = \mathrm{E}[(y_i - \mu_i)^2]$.

- Asymptotically can estimate $\sum_i \sigma_i^2 \mathbf{x}_i\mathbf{x}_i'$ by $\left(\sum_i (y_i - \widehat{\mu}_i)^2 \mathbf{x}_i\mathbf{x}_i'\right)$.
- If density is Poisson then simplifies to $\left(\sum_i \mu_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1}$ as $\sigma_i^2 = \mu_i$.

Poisson regression: data example

- Same 2003 MEPS data for over 65 in Medicare
- Dependent variable: `docvis`
- Regressors grouped into three categories:
  - Health insurance status indicators
    - `private`
    - `medicaid`
  - Socioeconomic
    - `age`
    - `age2`
    - `educyr`
  - Health status measures
    - `actlim`
    - `totchr`

Summary statistics

```
. describe docvis $xlist

              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
docvis          float  %9.0g                  # doctor visits
private         byte   %8.0g                  =1 if has private supplementary insuran
medicaid        byte   %8.0g                  =1 if has Medicaid public insurance
age             byte   %8.0g                  Age
age2            float  %9.0g                  Age-squared
educyr          byte   %8.0g                  Years of education
actlim          byte   %8.0g                  =1 if activity limitation
totchr          byte   %8.0g                  # chronic conditions

.
. summarize docvis $xlist, sep(10)

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      docvis |      3677    6.822682    7.394937          0        144
     private |      3677    .4966005    .5000564          0          1
    medicaid |      3677     .166712    .3727692          0          1
         age |      3677    74.24476    6.376638         65         90
        age2 |      3677    5552.936    958.9996       4225       8100
      educyr |      3677    11.18031    3.827676          0         17
      actlim |      3677     .333152    .4714045          0          1
      totchr |      3677    1.843351    1.350026          0          8
```


Poisson MLE with robust sandwich standard errors

```
. * Poisson with robust standard errors
. poisson docvis $xlist, vce(robust) nolog  // Poisson robust SEs

Poisson regression                                Number of obs   =       3677
                                                  Wald chi2(7)    =     720.43
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood =  -15019.64                 Pseudo R2       =     0.1297

------------------------------------------------------------------------------
             |               Robust
      docvis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     private |   .1422324    .036356     3.91   0.000      .070976    .2134889
    medicaid |   .0970005   .0568264     1.71   0.088    -.0143773    .2083783
         age |   .2936722   .0629776     4.66   0.000     .1702383    .4171061
        age2 |  -.0019311   .0004166    -4.64   0.000    -.0027475   -.0011147
      educyr |   .0295562   .0048454     6.10   0.000     .0200594     .039053
      actlim |   .1864213   .0396569     4.70   0.000     .1086953    .2641474
      totchr |   .2483898   .0125786    19.75   0.000     .2237361    .2730435
       _cons |  -10.18221   2.369212    -4.30   0.000    -14.82578   -5.538638
------------------------------------------------------------------------------
```

Poisson MLE with default ML standard errors
- these are misleadingly small

```
. * Poisson with default ML standard errors
. poisson docvis $xlist                          // Poisson default ML standard errors

Iteration 0:   log likelihood = -15019.656
Iteration 1:   log likelihood =  -15019.64
Iteration 2:   log likelihood =  -15019.64

Poisson regression                                Number of obs   =       3677
                                                  LR chi2(7)      =    4477.98
                                                  Prob > chi2     =     0.0000
Log likelihood =  -15019.64                       Pseudo R2       =     0.1297
```

docvis	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
private	.1422324	.0143311	9.92	0.000	.114144	.1703208
medicaid	.0970005	.0189307	5.12	0.000	.0598969	.134104
age	.2936722	.0259563	11.31	0.000	.2427988	.3445457
age2	-.0019311	.0001724	-11.20	0.000	-.0022691	-.0015931
educyr	.0295562	.001882	15.70	0.000	.0258676	.0332449
actlim	.1864213	.014566	12.80	0.000	.1578726	.2149701
totchr	.2483898	.0046447	53.48	0.000	.2392864	.2574933
_cons	-10.18221	.9720115	-10.48	0.000	-12.08732	-8.277101

Robust se's are 2.5-2.7 times larger

Note: $\sqrt{s_y^2/\bar{y}} = \sqrt{7.39^2/6.82} = \sqrt{8.01} = 2.830$.

Poisson regression: coefficient interpretation

For the exponential conditional mean the marginal effect

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

1. Conditional mean is strictly monotonic increasing (or decreasing) in x_{ij} according to the sign of β_j.
2. Coefficients are semi-elasticities: β_j is proportionate change in conditional mean when x_{ij} changes by one unit.
3. More precisely $(\exp(\beta_j) - 1)$ is proportionate change. Programs have options to report exponentiated coefficients (incidence-rate ratios).
4. Like all single-index models, if $\beta_j = 2\beta_k$, then the effect of one-unit change in x_j is twice that of x_k.

Poisson regression: marginal effects

- Example: $\widehat{\beta}_{\text{Private}} = 0.142$.
  - Private insurance is associated with an increase in mean doctor visits of 14.2%.
  - More precisely the increase is $100 \times (e^{0.142} - 1) = 100 \times (1.153 - 1) = 15.3\%$.
  - Alternatively the exponentiated coefficient is $e^{0.142} = 1.153$, so the multiplicative effect is 1.153.
- Example: $\widehat{\beta}_{\text{Private}} = 0.142$ and $\widehat{\beta}_{\text{totchr}} = 0.248$
  - Since $0.142/0.248 = 0.57$, private insurance has the same impact on mean doctor visits as 0.57 more chronic conditions.

- Marginal effect at mean (MEM): Evaluate at $\mathbf{x} = \bar{\mathbf{x}}$

$$\text{MEM} = \left.\frac{\partial \mathrm{E}[y|\mathbf{x}]}{\partial x_j}\right|_{\mathbf{x}=\bar{\mathbf{x}}} = \exp(\bar{\mathbf{x}}'\widehat{\boldsymbol{\beta}}) \times \widehat{\beta}_j$$

- Average marginal effect (AME): Evaluate at each $\mathbf{x}_i$ and average

$$\text{AME} = \sum\nolimits_i \frac{\partial \mathrm{E}[y_i|\mathbf{x}_i]}{\partial x_{ij}} = \sum\nolimits_i \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}) \times \widehat{\beta}_j.$$

- For Poisson with intercept in model AME $= \bar{y}\widehat{\beta}_j$
  - Reason: f.o.c. $\sum_i (y_i - \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})) = 0$ imply $\sum_i \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}) = \bar{y}$
- For Poisson can show that AME > MEM.

Marginal effects: MEM versus AME

```
. mfx      // MEM: Marginal effect for Poisson evaluated at average of x

Marginal effects after poisson
      y  = predicted number of events (predict)
         =  6.2674204
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
 private*|   .8926136      .22788    3.92   0.000   .445984  1.33924    .4966
medicaid*|   .6281643      .38182    1.65   0.100  -.120185  1.37651  .166712
     age |   1.840567      .39247    4.69   0.000   1.07134  2.60979  74.2448
    age2 |   -.012103       .0026   -4.66   0.000  -.017194 -.007012  5552.94
  educyr |   .1852412      .03067    6.04   0.000   .125127  .245355  11.1803
 actlim* |   1.207039      .26864    4.49   0.000   .680506  1.73357  .333152
  totchr |   1.556764      .07602   20.48   0.000   1.40777  1.70575  1.84335
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

. margeff   // AME: Average marginal effect for Poisson

Average partial effects after poisson
      y  = E(docvis) (expected number of counts)

------------------------------------------------------------------------------
    variable |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     private |   .9701906   .2645543     3.67   0.000     .4516738    1.488707
    medicaid |   .6830664   .4219606     1.62   0.105    -.1439612    1.510094
         age |   2.032557   .4489563     4.53   0.000     1.152618    2.912495
        age2 |  -.0131753   .0028473    -4.63   0.000    -.0187559   -.0075947
      educyr |    .201682    .033795     5.97   0.000      .135445    .2679189
      actlim |   1.295942    .304792     4.25   0.000     .6985604    1.893323
      totchr |   1.712165   .0935177    18.31   0.000     1.528874    1.895456
------------------------------------------------------------------------------
```


Generalized linear models

- 1. Specify the link function (links the mean to $\mathbf{x}_i'\beta$)

$$f(\mathrm{E}[y_i|\mathbf{x}_i]) = \mathbf{x}_i'\beta.$$

  - This implies the conditional mean is

$$\mathrm{E}[y_i|\mathbf{x}_i] = g(\mathbf{x}_i'\beta) = f^{-1}(\mathbf{x}_i'\beta).$$

  - So $g(\cdot)$ is the inverse link function.
- 2. Specify the conditional variance to be a multiple of a specified function $h(\cdot)$ of the mean

$$\mathrm{V}[y_i|\mathbf{x}_i] = \alpha \times h(\mathrm{E}[y_i|\mathbf{x}_i])$$

- 3. The GLM estimator solves

$$\sum_{i=1}^N \frac{(y_i - g(\mathbf{x}_i'\beta))}{h(g(\mathbf{x}_i'\beta))} \times g'(\mathbf{x}_i'\beta)\mathbf{x}_i = \mathbf{0}.$$

  - This is GLS-like: $\sum_i \frac{(y_i - \mathrm{E}[y_i|\mathbf{x}_i])}{\mathrm{V}[y_i|\mathbf{x}_i]} \times \frac{\partial \mathrm{E}[y_i|\mathbf{x}_i]}{\partial \mathbf{x}_i} = \mathbf{0}$.
  - Consistent if $\mathrm{E}[(y_i - g(\mathbf{x}_i'\beta)] = 0$ or $\mathrm{E}[y_i|\mathbf{x}_i] = g(\mathbf{x}_i'\beta)$.
  - Use sandwich s.e.'s, or rescale by $\sqrt{\widehat{\alpha}}$ where $\widehat{\alpha} = \frac{1}{n-k}\sum_i \frac{(y_i - \widehat{\mu}_i)^2}{h(\widehat{\mu}_i)}$

GLM: Poisson as GLM

- Poisson specifies a log-link function and variance = mean so

$$\begin{aligned}
\mathrm{E}[y_i|\mathbf{x}_i] &= g(\mathbf{x}_i'\beta) = \exp(\mathbf{x}_i'\beta) \\
\ln \mathrm{E}[y_i|\mathbf{x}_i] &= \mathbf{x}_i'\beta \\
h(\mathrm{E}[y_i|\mathbf{x}_i]) &= \exp(\mathbf{x}_i'\beta)
\end{aligned}$$

- Then GLM for Poisson solves.

$$\begin{aligned}
\sum_i \frac{(y_i - \exp(\mathbf{x}_i'\beta))}{\exp(\mathbf{x}_i'\beta)} \exp(\mathbf{x}_i'\beta)\mathbf{x}_i &= \mathbf{0} \\
\text{or } \sum_i (y_i - \exp(\mathbf{x}_i'\beta)) &= \mathbf{0}.
\end{aligned}$$

  - This is exactly same as Poisson MLE f.o.c.
  - Simplification occurred because the link here was the canonical link.

Poisson GLM with robust sandwich standard errors

```
. glm docvis $xlist, family(poisson) link(log) vce(robust) nolog

Generalized linear models                          No. of obs      =      3677
Optimization     : ML                              Residual df     =      3669
                                                   Scale parameter =         1
Deviance         =  18395.14033                    (1/df) Deviance =  5.013666
Pearson          =  23147.37781                    (1/df) Pearson  =  6.308906

Variance function: V(u) = u                        [Poisson]
Link function    : g(u) = ln(u)                    [Log]

                                                   AIC             =  8.173859
Log pseudolikelihood = -15019.6398                 BIC             = -11726.81

------------------------------------------------------------------------------
             |               Robust
      docvis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     private |   .1422324    .036356     3.91   0.000      .070976    .2134889
    medicaid |   .0970005   .0568264     1.71   0.088    -.0143773    .2083783
         age |   .2936722   .0629776     4.66   0.000     .1702383    .4171061
        age2 |  -.0019311   .0004166    -4.64   0.000    -.0027475   -.0011147
      educyr |   .0295562   .0048454     6.10   0.000     .0200594     .039053
      actlim |   .1864213   .0396569     4.70   0.000     .1086953    .2641474
      totchr |   .2483898   .0125786    19.75   0.000     .2237361    .2730435
       _cons |  -10.18221   2.369212    -4.30   0.000    -14.82578   -5.538638
------------------------------------------------------------------------------
```

Exactly same as `poisson, vce(robust)`

Poisson GLM with rescaled ML standard errors.
Multiply by $\sqrt{\widehat{\alpha}}$ where $\widehat{\alpha} = \frac{1}{n-k}\sum_i \frac{(y_i-\widehat{\mu}_i)^2}{h(\widehat{\mu}_i)}$ = `(1/df) Pearson` $= 6.308906$.

```
. glm docvis $xlist, family(poisson) link(log) scale(x2) nolog

Generalized linear models                          No. of obs      =      3677
Optimization     : ML                              Residual df     =      3669
                                                   Scale parameter =         1
Deviance         =  18395.14033                    (1/df) Deviance =  5.013666
Pearson          =  23147.37781                    (1/df) Pearson  =  6.308906

Variance function: V(u) = u                        [Poisson]
Link function    : g(u) = ln(u)                    [Log]

                                                   AIC             =  8.173859
Log likelihood   = -15019.6398                     BIC             = -11726.81

------------------------------------------------------------------------------
             |                 OIM
      docvis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     private |   .1422324   .0359961     3.95   0.000     .0716813    .2127836
    medicaid |   .0970005   .0475493     2.04   0.041     .0038055    .1901954
         age |   .2936722   .0651959     4.50   0.000     .1658906    .4214538
        age2 |  -.0019311   .0004331    -4.46   0.000     -.00278    -.0010822
      educyr |   .0295562   .0047271     6.25   0.000     .0202912    .0388212
      actlim |   .1864213   .0365861     5.10   0.000     .1147139    .2581288
      totchr |   .2483898   .0116663    21.29   0.000     .2255243    .2712554
       _cons |  -10.18221   2.441453    -4.17   0.000    -14.96737   -5.397048
------------------------------------------------------------------------------
(Standard errors scaled using square root of Pearson X2-based dispersion)
```

Within 15% of robust sandwich standard errors.

GLM: As quasi-MLE for linear exponential family

- Class of models based on linear exponential family (LEF):
  - normal, binomial, Bernoulli, gamma, exponential, Poisson.
- Specifically for the LEF

$$\begin{aligned} f(y_i|\mu_i) &= \exp\{a(\mu_i)+b(y_i)+c(\mu_i)y_i\} \\ \mathsf{E}[y_i] &= \mu_i = -a'(\mu_i)/c(\mu_i) \\ \mathsf{V}[y_i] &= 1/c(\mu_i) \end{aligned}$$

- Poisson: $a(\mu) = -\mu$; $c(\mu) = \ln\mu$ so
  - $a'(\mu) = -1$
    $c'(\mu) = 1/\mu$
  - $\mathsf{E}[y] = -(-1)/(1/\mu) = \mu$
    $\mathsf{V}[y] = 1/(1/\mu) = \mu$.

- Quasi-MLE maximizes

$$\ln L(\boldsymbol{\beta}) = \textstyle\sum_i \ln f(y_i|\mu_i(\boldsymbol{\beta})) = \sum_i\{a(\mu_i(\boldsymbol{\beta}))+b(y_i)+c(\mu_i(\boldsymbol{\beta}))y_i\}.$$

- F.O.C. are

$$\begin{aligned} & \textstyle\sum_i\{a'(\mu_i(\boldsymbol{\beta}))+c'(\mu_i(\boldsymbol{\beta}))y_i\} \times \frac{\partial\mu_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} &= \mathbf{0} \\ \Rightarrow\ & \textstyle\sum_i c'(\mu_i(\boldsymbol{\beta}) \times \{y_i - a'(\mu_i(\boldsymbol{\beta}))/c'(\mu_i(\boldsymbol{\beta}))\} \times \frac{\partial\mu_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} &= \mathbf{0} \\ \Rightarrow\ & \textstyle\sum_i \frac{1}{\mathsf{V}[y_i]}\{y_i - \mu_i(\boldsymbol{\beta})\} \times \frac{\partial\mu_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} &= \mathbf{0} \end{aligned}$$

- MLE based on LEF with $\mu_i = g(\mathbf{x}_i'\boldsymbol{\beta})$ shares the robustness properties of normal and Poisson MLE
  - consistency requires correct specification of the mean.
- But correct standard errors should use a robust estimate of variance
  - Robust sandwich s.e.'s or
  - Default ML s.e.'s multiplied by $\sqrt{\widehat{\alpha}}$ where $\mathsf{V}[y_i|\mathbf{x}_i] = \alpha \times h(\mathsf{E}[y_i|\mathbf{x}_i])$.

Nonlinear least squares estimator

- Alternative estimator not used as Poisson is simpler and usually more efficient.
- Specify same conditional mean as Poisson: $\mathsf{E}[y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta})$.
- Minimize sum of squared residuals: $\sum_{i=1}^N (y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))^2$.
- NLS first-order conditions:

$$\textstyle\sum_{i=1}^N (y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})\mathbf{x}_i = \mathbf{0}.$$

- NLS is consistent provided $\mathsf{E}[y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta})$
- $\widehat{\boldsymbol{\beta}}_{\mathsf{NLS}} \overset{a}{\sim} \mathcal{N}[\boldsymbol{\beta}, \mathsf{V}_{\mathsf{MLE}}[\widehat{\boldsymbol{\beta}}]]$ and use robust sandwich variance estimate:

$$\widehat{\mathsf{V}}_{\mathsf{ROB}}[\widehat{\boldsymbol{\beta}}_{\mathsf{NLS}}] = \left(\textstyle\sum_i \widehat{\mu}_i^2\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\left(\sum_i (y_i-\widehat{\mu}_i)^2\widehat{\mu}_i^2\mathbf{x}_i\mathbf{x}_i'\right)\left(\sum_i \widehat{\mu}_i^2\mathbf{x}_i\mathbf{x}_i'\right)^{-1}.$$

Nonlinear least squares with robust sandwich standard errors

```
. nl (docvis = exp({xb: $xlist one})), vce(robust) nolog
(obs = 3677)

Nonlinear regression                              Number of obs =       3677
                                                  R-squared     =     0.5436
                                                  Adj R-squared =     0.5426
                                                  Root MSE      =   6.804007
                                                  Res. dev.     =   24528.25

------------------------------------------------------------------------------
             |               Robust
      docvis |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 /xb_private |   .1235144   .0395179     3.13   0.002     .0460351    .2009937
/xb_medicaid |   .0856747   .0649936     1.32   0.188    -.0417525    .2131018
     /xb_age |   .2951153   .0720509     4.10   0.000     .1538516    .4363789
    /xb_age2 |  -.0019481   .0004771    -4.08   0.000    -.0028836   -.0010127
  /xb_educyr |   .0309924   .0051192     6.05   0.000     .0209557    .0410291
  /xb_actlim |   .1916735   .0413705     4.63   0.000      .110562    .2727851
  /xb_totchr |   .2191967   .0151021    14.51   0.000     .1895874      .24886
     /xb_one |  -10.12438   2.713159    -3.73   0.000    -15.44383   -4.804931
------------------------------------------------------------------------------
```

Robust standard errors are 5-20% larger than those for Poisson.

Diagnostics: residuals and influence measures

- Residuals (for Poisson)
  - Raw: $r_i = (y_i - \widehat{\mu}_i)$
  - Pearson: $p_i = (y_i - \widehat{\mu}_i)/\sqrt{\widehat{\mu}_i}$
  - Deviance: $d_i = sign(y_i - \widehat{\mu}_i)\sqrt{2\{y_i \ln(y_i/\widehat{\mu}_i) - (y_i - \widehat{\mu}_i)\}}$
  - Anscombe: $a_i = 1.5(y_i^{2/3} - \mu_i^{2/3})/\mu_i^{1/6}$
  - Last three will be standardized if $\mathsf{V}[y_i] = \mu_i$.
- Small-sample corrections (for Poisson)
  - Hat matrix: $\mathbf{H} = \mathbf{W}^{1/2}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{1/2}$; $\mathbf{W} = \text{Diag}[\widehat{\mu}_i]$.
  - Studentized residual: $p_i^* = p_i/\sqrt{1-h_{ii}}$ and $d_i^* = d_i/\sqrt{1-h_{ii}}$.
- Influential observations:
  - Rule of thumb: $h_{ii} > 2K/N$
  - Cook's distance: $C_i = (p_i^*)^2 h_i/K(1-h_{ii})$ measures change in $\widehat{\beta}$ when observation i is omitted.


```
. summarize rraw rpearson rdeviance ranscombe hat cooksd, sep(10)

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        rraw |      3677   -7.08e-10    6.808178  -29.12996   136.9007
    rpearson |      3677   -.0060737    2.509354  -4.914746   51.38051
   rdeviance |      3677   -.3438453    2.210397  -6.087067   24.35213
   ranscombe |      3677   -.3634087    2.254119  -6.153762   25.72892
         hat |      3677    .0021757    .0015322   .0007023    .027556
      cooksd |      3677    .0019357    .0177516   6.54e-11    .964198

.
. correlate rraw rpearson rdeviance ranscombe
(obs=3677)

             |     rraw rpearson rdevia~e ransco~e
-------------+------------------------------------
        rraw |   1.0000
    rpearson |   0.9792   1.0000
   rdeviance |   0.9454   0.9669   1.0000
   ranscombe |   0.9435   0.9661   0.9998   1.0000
```

The various residuals are highly correlated.
The raw residuals sum to zero due to f.o.c.

Diagnostics: R-squared measures

- Different interpretations of R^2 in linear model lead to different R^2 in nonlinear model. Most are difficult to interpret in nonlinear models.
- Simplest: squared correlation coefficient between y_i and $\widehat{y}_i = \widehat{\mu}_i$

$$R^2_{\text{Cor}} = \widehat{\text{Cor}}^2[y_i, \widehat{y}_i]$$

- Sums of squares measures differ in nonlinear models

$$R^2_{\text{Res}} = 1 - \text{ResSS}/\text{TotalSS}$$
$$R^2_{\text{Exp}} = \text{ExpSS}/\text{TotalSS}$$

- Relative gain in log-likelihood (L_0 is intercept model only)

$$R^2_{\text{RG}} = \frac{\ln L_{fit} - \ln L_0}{\ln L_{\max} - \ln L_0} = 1 - \frac{\ln L_{\max} - \ln L_{fit}}{\ln L_{\max} - \ln L_0}.$$

  - Works for Poisson as $\ln L_{\max}$ occurs when $\mu_i = y_i$.
  - Unlike others $0 \leq R^2_{\text{RG}} < 1$ and R^2_{RG} always increases as add regressors.
- Stata measure is only applicable to binary and multinomial models

$$R^2_{\text{Pseudo}} = 1 - \ln L_{fit}/\ln L_0.$$

Diagnostics: overdispersion test

- H_0 :$\mathrm{V}[y_i|\mathbf{x}_i] = \mathrm{E}[y_i|\mathbf{x}_i]$ versus H_1 :$\mathrm{V}[y_i|\mathbf{x}_i] = \mathrm{E}[y_i|\mathbf{x}_i] + \alpha(\mathrm{E}[y_i|\mathbf{x}_i])^2$.
  Test $H_0 : \alpha = 0$ against $H_1 : \alpha > 0$.
- Implement by auxiliary regression

$$((y_i - \widehat{\mu}_i)^2 - y_i)/\widehat{\mu}_i = \alpha\widehat{\mu}_i + \ error$$

  and do t test of whether the coefficient of $\widehat{\mu}_i$ is zero.
- In practice can skip this test and just do Poisson with robust s.e.'s.
- Test is useful as can also use this test for test of underdispersion, whereas other test such Poisson versus negative binomial only test overdispersion.

Example of overdispersion test.

```
. * Overdispersion test against V[y|x] = E[y|x] + a*(E[y|x]^2)
. quietly poisson docvis $xlist, vce(robust)

. predict muhat, n

. quietly generate ystar = ((docvis-muhat)^2 - docvis)/muhat

. regress ystar muhat, noconstant noheader
```

ystar	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
muhat	.7047319	.1035926	6.80	0.000	.5016273	.9078365

Very strongly reject H_0. Data here are overdispersed.

Diagnostics: predicted probabilities

- Observed frequency $\overline{p}_j$ (fraction of observations with $y_i = j$).
- Fitted frequency $\widehat{p}_j = N^{-1}\sum_{i=1}^{N}\widehat{p}_{ij}$
  - predicted probability $\widehat{p}_{ij} = \Pr[y_i = j] = e^{-\widehat{\mu}_i}\widehat{\mu}_i^j/j!$ for Poisson.
- Expect $\widehat{p}_j$ close to $\bar{p}_j$, $j = 0, 1, 2, \ldots$.
- Informal statistic is Pearson's chi-square test

$$\sum_j \frac{(n\overline{p}_j - n\widehat{p}_j)^2}{n\widehat{p}_j}$$

  but this is not χ^2 distributed due to estimation to get $\widehat{p}_j$.

- Instead do a formal chi-square goodness of fit test.
- Assuming that the density is correctly specified (so more applicable to models more general than Poisson) this can be computed as NR_u^2 (uncentered R^2) from the artificial regression

$$1 = \mathbf{s}_i(y_i, \mathbf{x}_i, \widehat{\boldsymbol{\theta}})'\gamma + \sum_j (d_{ij}(y_i) - \widehat{p}_{ij})'\delta_j + error$$

- where
  - j denotes cells (e.g. values 0, 1, 2, 3, and 4 or more)
  - $d_{ij}(y_i)$ equals 1 if y_i is in cell j and 0 otherwise
  - $\widehat{p}_{ij}$ equals predicted probability for that cell
  - $\mathbf{s}_i(y_i, \mathbf{x}_i, \boldsymbol{\theta}) = \partial \ln f(y_i|\mathbf{x}_i, \boldsymbol{\theta})/\partial\boldsymbol{\theta}$ $\quad (= (y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))$ for Poisson).
- Reject at level α if $NR_u^2 > \chi_\alpha^2(J-1)$ where J is number of cells.

Compare actual and fitted frequencies for Poisson.

```
. countfit docvis $xlist, maxcount(10) prm nograph noestimates nofit
Comparison of Mean Observed and Predicted Count

             Maximum          At        Mean
Model      Difference      Value       |Diff|
---------------------------------------------
PRM          0.102            0         0.040

PRM: Predicted and actual probabilities

Count   Actual     Predicted     |Diff|    Pearson
------------------------------------------------
0        0.109       0.007        0.102   5168.233
1        0.085       0.030        0.056    387.868
2        0.097       0.063        0.034     69.000
3        0.091       0.095        0.005      0.789
4        0.092       0.116        0.024     17.861
5        0.072       0.121        0.049     72.441
6        0.063       0.114        0.051     84.355
7        0.055       0.099        0.044     73.016
8        0.049       0.082        0.033     50.128
9        0.042       0.065        0.024     31.225
10       0.029       0.051        0.021     33.402
------------------------------------------------
Sum      0.785       0.844        0.443   5988.318
```

Clearly problem as Poisson greatly underpredicts low counts e.g. for $y = 0$.

Negative binomial regression

- Negative binomial (Negbin 2) permits overdispersion.

$$f(y|\lambda,\alpha) = \frac{\Gamma(y+\alpha^{-1})}{\Gamma(y+1)\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\lambda}\right)^{\alpha^{-1}}\left(\frac{\lambda}{\alpha^{-1}+\lambda}\right)^{y}.$$

- Same conditional mean but different conditional variance to Poisson

$$\begin{aligned} \mathsf{E}[y|\mathbf{x}] &= \lambda = \exp(\mathbf{x}'\boldsymbol{\beta}) \\ \mathsf{V}[y|\mathbf{x}] &= \lambda + \alpha\lambda^2 = \exp(\mathbf{x}'\boldsymbol{\beta}) + \alpha(\exp(\mathbf{x}'\boldsymbol{\beta}))^2. \end{aligned}$$

- The ML first-order conditions w.r.t. $\boldsymbol{\beta}$ and α are (with $\mu_i = \exp(\mathbf{x}_i'\boldsymbol{\beta})$)

$$\sum_{i=1}^N \frac{y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta})}{1+\alpha\exp(\mathbf{x}_i'\boldsymbol{\beta})}\mathbf{x}_i = \mathbf{0}$$

$$\sum_{i=1}^N \left\{\frac{1}{\alpha^2}\left(\ln(1+\alpha\mu_i) - \sum_{j=0}^{y_i-1}\frac{1}{(j+\alpha^{-1})}\right) + \frac{y_i-\mu_i}{\alpha(1+\alpha\mu_i)}\right\} = 0.$$

- Can additionally allow $\alpha = \exp(\mathbf{x}'\gamma)$ (generalized negative binomial).
- Can instead use Negbin 1: $\mathsf{V}[y|\mathbf{x}] = (1+\alpha)\lambda = (1+\alpha)\exp(\mathbf{x}'\boldsymbol{\beta})$.
- Often little efficiency gain (if any) over Poisson with robust s.e's.

Negative binomial MLE with ML default standard errors

```
. nbreg docvis $xlist, nolog

Negative binomial regression                      Number of obs   =       3677
                                                  LR chi2(7)      =     773.44
Dispersion     = mean                             Prob > chi2     =     0.0000
Log likelihood = -10589.339                       Pseudo R2       =     0.0352

------------------------------------------------------------------------------
      docvis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     private |   .1640928   .0332186     4.94   0.000     .0989856    .2292001
    medicaid |    .100337   .0454209     2.21   0.027     .0113137    .1893603
         age |   .2941294   .0601588     4.89   0.000     .1762203    .4120384
        age2 |  -.0019282   .0004004    -4.82   0.000    -.0027129   -.0011434
      educyr |   .0286947   .0042241     6.79   0.000     .0204157    .0369737
       actlim |   .1895376   .0347601     5.45   0.000      .121409    .2576662
      totchr |   .2776441   .0121463    22.86   0.000     .2538378    .3014505
       _cons |  -10.29749   2.247436    -4.58   0.000    -14.70238   -5.892595
-------------+----------------------------------------------------------------
    /lnalpha |  -.4452773   .0306758                     -.5054007   -.3851539
-------------+----------------------------------------------------------------
       alpha |   .6406466   .0196523                      .6032638    .6803459
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) = 8860.60 Prob>=chibar2 = 0.000
```

Likelihood ratio test of $\alpha = 0$ prefers NB to Poisson ($p < 0.05$)
- where critical values use half $\chi^2(1)$ as $\alpha = 0$ is on boundary of NB.
Fitted frequencies close to observed frequencies (from output not given).

Poisson and negative binomial MLE with different standard error estimates

	(1) PDEFAULT	(2) PROBUST	(3) PPEARSON	(4) NBDEFAULT	(5) NBROBUST
docvis					
private	0.1422***	0.1422***	0.1422***	0.1641***	0.1641***
	(0.0143)	(0.0364)	(0.0360)	(0.0332)	(0.0369)
medicaid	0.0970***	0.0970	0.0970*	0.1003*	0.1003
	(0.0189)	(0.0568)	(0.0475)	(0.0454)	(0.0567)
age	0.2937***	0.2937***	0.2937***	0.2941***	0.2941***
	(0.0260)	(0.0630)	(0.0652)	(0.0602)	(0.0646)
age2	-0.0019***	-0.0019***	-0.0019***	-0.0019***	-0.0019***
	(0.0002)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
educyr	0.0296***	0.0296***	0.0296***	0.0287***	0.0287***
	(0.0019)	(0.0048)	(0.0047)	(0.0042)	(0.0049)
actlim	0.1864***	0.1864***	0.1864***	0.1895***	0.1895***
	(0.0146)	(0.0397)	(0.0366)	(0.0348)	(0.0394)
totchr	0.2484***	0.2484***	0.2484***	0.2776***	0.2776***
	(0.0046)	(0.0126)	(0.0117)	(0.0121)	(0.0132)
_cons	-10.1822***	-10.1822***	-10.1822***	-10.2975***	-10.2975***
	(0.9720)	(2.3692)	(2.4415)	(2.2474)	(2.4241)
lnalpha					
_cons				-0.4453***	-0.4453***
				(0.0307)	(0.0378)
N	3677	3677	3677	3677	3677
pseudo R-sq	0.130	0.130		0.035	0.035

standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Summary

- Poisson regression (or GLM) is straightforward
  - many packages do Poisson regression
  - coefficients are easily interpreted as semi-elasticities.
- Do Poisson rather than OLS with dependent variable
  - y; $\ln y$ (with adjustment for $\ln 0$); or $\sqrt{y}$.
- Poisson MLE is consistent provided only that $\mathsf{E}[y|\mathbf{x}] = \exp(\mathbf{x}'\boldsymbol{\beta})$.
  - But make sure standard errors etc. are robust to $\mathsf{V}[y|\mathbf{x}] \neq \mathsf{E}[y|\mathbf{x}]$.
- But if need to predict probabilities use a richer model.
  - Good starting point is negative binomial
  - Additional models are discussed next.

References

This material is in standard texts.

- Count data models in addition to Cameron and Trivedi books:
  - Winkelmann, R. (2008), *Econometric Analysis of Count Data*, 5th edition, Springer.
  - Hilbe, J. (2007), *Negative Binomial Regression*, Cambridge University Press.
  - Cameron, A.C., and P.K. Trivedi (1986), "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators," *Journal of Applied Econometrics*, 1, 29-53.
- Generalized linear models books:
  - McCullagh, P. and J.A. Nelder (1989), *Generalized Linear Models*, Second Edition, Chapman and Hall.
  - Dobson, A.J. and A. Barnett (2008), *An Introduction to Generalized Linear Models*, Third Edition, Chapman and Hall.
  - Hardin, J.W. and J.M. Hilbe (2007), *Generalized Linear Models and Extensions*, Second Edition, Stata Press.