Monday Part 1
Counts: Cross-section Basics

A. Colin Cameron
Univ. of Calif. - Davis

. .

Nonlinear Cross-section and Panel Regression Models for Count Data

. .

Queens University, Canada
Department of Economics

May 13-16, 2019

1. 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).

- Analysis is straightforward in usual case of model the conditional mean $\mathrm{E}[y|\mathbf{x}]$:
  - in Stata replace command `regress` with `poisson`
  - and replace command `xtreg` with command `xtpoisson`
- Interpretation of marginal effects, however, is more complicated:
  - $\mathrm{E}[y|\mathbf{x}] = \exp(\mathbf{x}'\boldsymbol{\beta})$ so $\mathrm{ME}_j = \partial \mathrm{E}[y|\mathbf{x}]/\partial x_j = \beta_j \exp(\mathbf{x}'\boldsymbol{\beta}) \neq \beta_j$.
- Analysis is more complicated for
  - better parametric models for prediction, censoring, selection
  - time series of counts.

Outline

2. Poisson - Data Example: Doctor visits (MEPS)

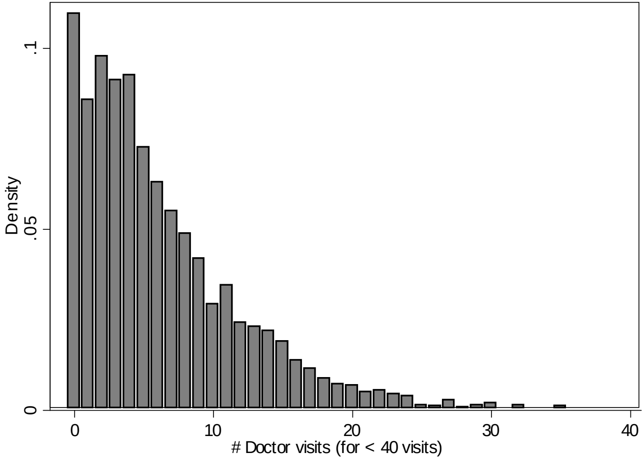
- Many health surveys measure health use as counts as people have better recall of counts than of dollars spent.
- 2003 U.S. Medical Expenditure Panel Survey (MEPS).
- Sample of Medicare population aged 65 and higher (N = 3,677)
- docvis = annual number of doctor visits

```
. 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



Poisson distribution

- From stochastic process theory, natural model for counts is

$$y \sim Poisson[\lambda].$$

- Probability mass function:

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

- Mean and variance:

$$\mathrm{E}[y] = \lambda \quad \text{and} \quad \mathrm{V}[y] = \lambda$$

- Equidispersion: variance = mean
  - Restriction imposed by Poisson
- Overdispersion: variance > mean
  - More common feature of count data
  - Doctor visits data: $\bar{y} = 6.82$, $s_y^2 = 54.68 \simeq 8.01\bar{y}$.

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 $\lambda > 0$.
$$\lambda = \mathsf{E}[y|\mathbf{x}] = \exp(\mathbf{x}'\boldsymbol{\beta}).$$
- MLE is straightforward given data independent over i.

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

Poisson regression: first-order conditions

- 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 $\boldsymbol{\beta}$
  - $\widehat{\boldsymbol{\beta}}_{s+1} - \widehat{\boldsymbol{\beta}}_s = -\mathbf{H}_s^{-1}\mathbf{g}_s$ where
  - $\mathbf{H}_s = \partial^2 \ln L(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'\big|_{\widehat{\boldsymbol{\beta}}_s} = -\sum_{i=1}^{n}\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_s)\mathbf{x}_i\mathbf{x}_i'$
  - and $\mathbf{g}_s = \partial \ln L(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\big|_{\widehat{\boldsymbol{\beta}}_s} = \sum_{i=1}^{n}(y_i - \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}_s))\mathbf{x}_i$.

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 (OLS).

Poisson regression: distribution of Poisson MLE

- In general $\widehat{\beta} \overset{a}{\sim} \mathcal{N}[\beta, \mathrm{V}[\widehat{\beta}]]$
- Given independence over i there are three different estimates of $\mathrm{V}[\widehat{\beta}]$
  - 1. Do not use: Assume $y_i|\mathbf{x}_i$ distribution is Poisson. Then

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

  - 2. Econometrics: Relax Poisson assumption though assume $\mathrm{E}[y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\beta)$. Then use the heteroskedastic-robust estimate

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

  - 3. Statistics GLM: Relax Poisson assumptions though assume $\mathrm{E}[y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\beta)$ and $\mathrm{V}[y_i|\mathbf{x}_i] = \alpha \times \exp(\mathbf{x}_i'\beta)$. Then

$$\widehat{\mathrm{V}}_{\mathrm{GLM}}[\widehat{\beta}] = \widehat{\alpha} \times \left(\sum_i \widehat{\mu}_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1}, \ \widehat{\alpha} = \frac{1}{n-k} \sum_{i=1}^{n} (y_i - \widehat{\mu}_i)^2.$$

- Typically $\widehat{\mathrm{V}}_{\mathrm{HET}}[\widehat{\beta}] \simeq \widehat{\mathrm{V}}_{\mathrm{GLM}}[\widehat{\beta}] >> \widehat{\mathrm{V}}_{\mathrm{MLE}}[\widehat{\beta}]$.

Poisson regression: data example

- 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`
- `global xlist private medicaid age age2 educyr actlim totchr`
  - in commands refer to as `$xlist`

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 - preferred

```
. * 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
```

docvis	Coef.	Robust 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 - do not use
- These are misleadingly small due to overdispersion!!

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

3. Coefficient interpretation and marginal effects

- For the exponential conditional mean the marginal effect

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

- **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: coefficient interpretation (continued)

- 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.
  - In Stata `poisson y x, irr`
- 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 effects: Three types

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

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

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

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

- 3. Marginal effect at representative value (MER): Evaluate at $\mathbf{x} = \mathbf{x}^*$
- AME is nest
  - For population AME use population weights in computing AME.

- 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.
- Computation of marginal effects in Stata from Stata 11 onwards
  - after `poisson` (or other regression command) give command
  - `margins, dydx(*)` for AME
  - `margins, dydx(*) atmean` for MEM
  - `margins, dydx(*) at(age=30 educyr=12)` for MER

Marginal effects: AME (This page) versus MEM (next page)

```
. * AME and MEM for Poisson
. quietly poisson docvis $xlist, vce(robust)

. margins, dydx(*)          // AME: Average marginal effect for Poisson

Average marginal effects                        Number of obs     =      3,677
Model VCE    : Robust

Expression   : Predicted number of events, predict()
dy/dx w.r.t. : private medicaid age age2 educyr actlim totchr

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     private |   .9704067    .247564     3.92   0.000     .4851902    1.455623
    medicaid |   .6618034   .3901692     1.70   0.090    -.1029143    1.426521
         age |   2.003632   .4303207     4.66   0.000     1.160219    2.847045
        age2 |  -.0131753   .0028473    -4.63   0.000    -.0187559   -.0075947
      educyr |   .2016526   .0337805     5.97   0.000     .1354441    .2678612
      actlim |   1.271893   .2749286     4.63   0.000     .7330432    1.810744
      totchr |   1.694685   .0908884    18.65   0.000     1.516547    1.872823
------------------------------------------------------------------------------
```

Marginal effects: MEM

```
. margins, dydx(*) atmean // MEM: ME for Poisson evaluated at average of x

Conditional marginal effects                    Number of obs     =      3,677
Model VCE    : Robust

Expression   : Predicted number of events, predict()
dy/dx w.r.t. : private medicaid age age2 educyr actlim totchr
at           : private         =    .4966005 (mean)
               medicaid        =     .166712 (mean)
               age             =    74.24476 (mean)
               age2            =    5552.936 (mean)
               educyr          =    11.18031 (mean)
               actlim          =     .333152 (mean)
               totchr          =    1.843351 (mean)
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
private	.8914309	.2270816	3.93	0.000	.4463591	1.336503
medicaid	.607943	.3577377	1.70	0.089	-.09321	1.309096
age	1.840568	.3924682	4.69	0.000	1.071345	2.609792
age2	-.012103	.0025973	-4.66	0.000	-.0171936	-.0070125
educyr	.1852413	.0306709	6.04	0.000	.1251275	.2453551
actlim	1.168381	.2516143	4.64	0.000	.6752264	1.661536
totchr	1.556764	.0760166	20.48	0.000	1.407775	1.705754

Finite Differences

- The preceding used calculus to get marginal effects.
- For indicator variables such as for gender it makes more sense to consider a one unit change.
- If $\mathrm{E}[y|\mathbf{z}, d] = \exp(\mathbf{z}'\delta + \alpha d)$ then

$$\begin{aligned} \mathrm{ME}_{fd} &= \mathrm{E}[y|\mathbf{z}, d=1] - \mathrm{E}[y|\mathbf{z}, d=0] \\ &= \exp(\mathbf{z}'\delta + \alpha) - \exp(\mathbf{z}'\delta) \\ &= \exp(\mathbf{z}'\delta) \times (\exp(\alpha) - 1). \end{aligned}$$

- Finite difference marginal effects can be calculated using **factor variables**
  - these can also be used to get ME's for interactions
  - e.g. for age where age appears as a quadratic.

Factor variables

- Factor variables allow one to compute ME using finite differences and allowing interactions.
  - use `i.` for discrete variables and finite difference ME's
  - use `c.` for continuous variables and calculus ME's
  - and use # and ## for interactions.

```
. poisson docvis i.private i.medicaid c.age##c.age educyr i.actlim totchr, vce(robust)

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

Poisson regression                              Number of obs     =      3,677
                                                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]
-------------+----------------------------------------------------------------
   1.private |   .1422324    .036356     3.91   0.000      .070976    .2134889
  1.medicaid |   .0970005   .0568264     1.71   0.088    -.0143773    .2083783
         age |   .2936722   .0629776     4.66   0.000     .1702383    .4171061
             |
 c.age#c.age |  -.0019311   .0004166    -4.64   0.000    -.0027475   -.0011147
             |
      educyr |   .0295562   .0048454     6.10   0.000     .0200594     .039053
    1.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
------------------------------------------------------------------------------
```

`margins, dydx(*)` uses finite difference for `i.` and calculus for `c.`
Plus now have ME with respect to age allowing for quadratic.

```
. * Also factor variables and noncaculus methods
. * The i. are discrete and will calculate ME of one unit change (not derivative)
. * The c.age##age means age and age-sauared appear and ME is w.r.t. age
. quietly poisson docvis i.private i.medicaid c.age##c.age educyr i.actlim totchr, vce(robust)

. margins, dydx(*) // MEM: ME for Poisson evaluated at average of x

Average marginal effects                        Number of obs     =      3,677
Model VCE    : Robust

Expression   : Predicted number of events, predict()
dy/dx w.r.t. : 1.private 1.medicaid age educyr 1.actlim totchr

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   1.private |   .9701906   .2473149     3.92   0.000     .4854622    1.454919
  1.medicaid |   .6830664   .4153252     1.64   0.100     -.130956    1.497089
         age |   .0385842   .0172075     2.24   0.025     .0048581    .0723103
      educyr |   .2016526   .0337805     5.97   0.000     .1354441    .2678612
    1.actlim |   1.295942   .2850588     4.55   0.000     .7372367    1.854647
      totchr |   1.694685   .0908883    18.65   0.000     1.516547    1.872823
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

4. Generalized linear models

- Generalized linear models (GLM) is the framework used in the statistics literature for nonlinear regression.
- Leading examples are
  - OLS regression for $y \in (-\infty, \infty)$
  - Logit and probit regression for $y \in \{0, 1\}$
  - Poisson regression for $y \in \{0, 1, 2, 3, ...\}$
  - Gamma regression including exponential for $y \in (0, \infty)$
- In Stata use command `glm`
  - specify the GLM family member: here `poisson`
  - specify the link function (inverse of the conditional mean function): here `log`
  - get heteroskedastic-robust standard errors: `vce(robust)`

Poisson GLM with heteroskedastic-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
```

docvis	Coef.	Robust 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)`

ASIDE: What is a generalized linear model?

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

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

- For regression specify a model of the mean
  - $\mu_i = \mu_i(\boldsymbol{\beta}) = \mu_i(\mathbf{x}_i, \boldsymbol{\beta})$.
- Poisson is a member with
  - $a(\mu) = -\mu$; $c(\mu) = \ln \mu$
  - $a'(\mu) = -1$ and $c'(\mu) = 1/\mu$
  - $\mathrm{E}[y] = -(-1)/(1/\mu) = \mu$ and $\mathrm{V}[y] = 1/(1/\mu) = \mu$.

- Quasi-MLE for LEF maximizes the log-likelihood

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

- The first-order conditions are

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

- 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 (so $\mathsf{E}[\{y_i - \mu_i(\boldsymbol{\beta})\}] = 0$).
- But correct standard errors should use a robust estimate of variance
  - Robust sandwich standard errors or
  - Default ML s.e.'s multiplied by $\sqrt{\widehat{\alpha}}$ where $\mathsf{V}[y_i|\mathbf{x}_i] = \alpha \times (1/c'(\mu_i))$.

ASIDE: Nonlinear least squares estimator

- Alternative estimator (also consistent in the Poisson model)
  - not used because Poisson is simpler and is 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 is consistent provided $\mathsf{E}[y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta})$
- $\widehat{\boldsymbol{\beta}}_{\text{NLS}} \overset{a}{\sim} \mathcal{N}[\boldsymbol{\beta}, \mathsf{V}_{\text{MLE}}[\widehat{\boldsymbol{\beta}}]]$ and use het-robust variance estimate:

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

- For doctor visits data
  - `nl (docvis = exp({xb:  $xlist}+{b0})), vce(robust)`
  - NLS robust standard errors are 5-20% larger than those for Poisson
- Note: Linear OLS estimates are close to the in-sample AMEs for Poisson.

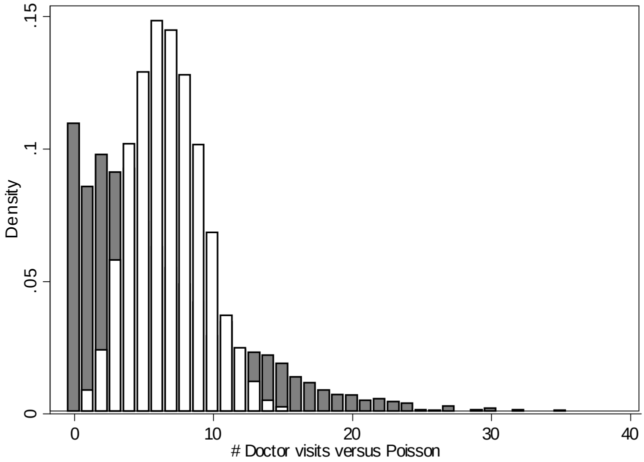
5. Negative binomial regression: motivation

- Count data are often overdispersed with more zeros and more high values than a Poisson distribution predicts.
- 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 (white) with $\lambda = \bar{y}$ compared to actual data (grey)



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

Negative binomial distribution

- Negative binomial is a Poisson-gamma mixture.

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

  then

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

- Probability mass function:

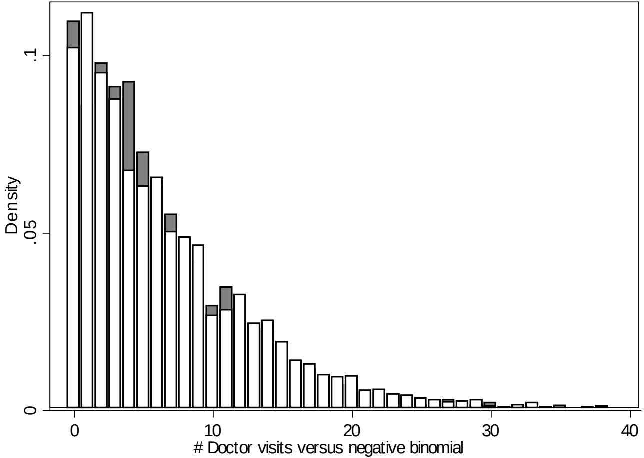
$$\Pr[Y = y|\lambda, \alpha] = \frac{\Gamma(\alpha^{-1} + y)}{\Gamma(\alpha^{-1})\Gamma(y+1)} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda}\right)^{\alpha^{-1}} \left(\frac{\lambda}{\lambda + \alpha^{-1}}\right)^{y}.$$

- 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.

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

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

- 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: $\mathrm{V}[y|\mathbf{x}]=(1+\alpha)\lambda=(1+\alpha)\exp(\mathbf{x}'\boldsymbol{\beta})$.
- 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
  - compare $\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}[y_i = k]$ to $\frac{1}{n}\sum_{i=1}^{n}\widehat{\Pr}[y_i = k]$ for $k = 0, 1, 2, \ldots$
- Use Stata add-on `chi2gof` which rejects model.

```
. chi2gof, cells(11) table

Chi-square Goodness-of-Fit Test for NegBin Model:

        Chi-square chi2(10) =    53.62
        Prob>chi2           =     0.00
```

Cells	Abs. Freq.	Rel. Freq.	Fitted Rel. Freq.	Abs. Dif.
0	401	.1091	.0913	.0178
1	314	.0854	.1079	.0225
2	358	.0974	.1054	.0081
3	334	.0908	.0962	.0053
4	339	.0922	.0849	.0073
5	266	.0723	.0735	.0012
6	231	.0628	.0631	2.6e-04
7	202	.0549	.0538	.0011
8	179	.0487	.0457	.0029
9	154	.0419	.0388	.003
10 or more	791	.2445	.2393	.0052

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

6. Summary of basic cross-section regression

- 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 later.

7. References

- A. Colin Cameron and Pravin K. Trivedi (2013)
  Regression Analysis of Count Data (RACD)
  Second edition, Cambridge Univ. Press.
- A. Colin Cameron and Pravin K. Trivedi (2005)
  Microeconometrics: Methods and Applications (MMA), chapter 20, Cambridge Univ. Press.
- A. Colin Cameron and Pravin K. Trivedi (2010)
  Microeconometrics using Stata (MUS), revised first edition, chapter 17, Stata Press.

Further References

- Count data models in addition to Cameron and Trivedi books:
  - Winkelmann, R. (2008), *Econometric Analysis of Count Data*, 5th edition, Springer.
  - Hilbe, J. (2011), *Negative Binomial Regression*, Second Edition, Cambridge University Press.
  - Hilbe, J. (2014), *Modeling Count Data*, 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 (2018), *An Introduction to Generalized Linear Models*, Fourth Edition, Chapman and Hall.
  - Hardin, J.W. and J.M. Hilbe (2012), *Generalized Linear Models and Extensions*, Third Edition, Stata Press.