Day 1A
Count Data Regression: Part 1

A. Colin Cameron
Univ. of Calif. - Davis
... for
Center of Labor Economics
Norwegian School of Economics
Advanced Microeconometrics

Aug 28 - Sept 1, 2017

1. Introduction

- We consider **nonlinear regression with cross-section data.**
- The slides use **count regression as a great illustrative example.**
- Count data models are for dependent variable $y = 0, 1, 2, \ldots$

- 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:
    - Fully parametric: MLE
    - Conditional mean: Quasi-MLE, generalized methods of moments (GMM)
  - Statistics: generalized linear models (GLM).

- Analysis is straightforward in the usual case of model the conditional mean $\mathrm{E}[y|\mathbf{x}]$:
  - in Stata replace command `regress` with `poisson`
  - and for panel data (later) 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
  - autoregressive time series of counts (not covered here).\
- **The session will focus on topics 2, 3 and 4.**

Outline

2. Poisson cross–section regression theory

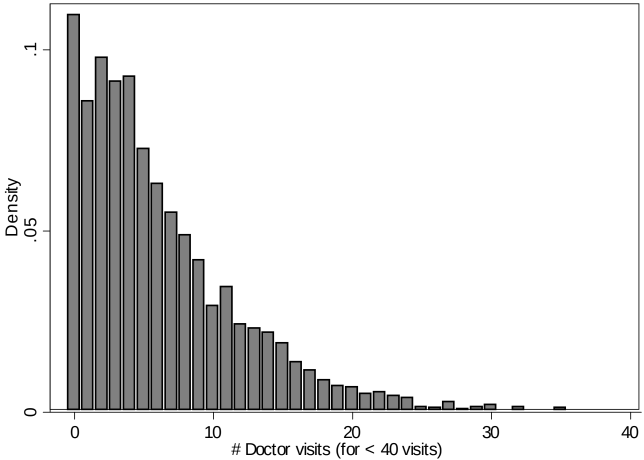
- Count data example
  - Many health surveys measure health use as counts as people have better recall of counts than of dollars spent.
  - 2003 U.S. Medical Expenditure Panel Survey (MEPS).
  - Sample of Medicare population aged 65 and higher (N = 3,677)
  - docvis = annual number of doctor visits

```
. 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:

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

- 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 $\mathsf{E}[y|\mathbf{x}] = \exp(\mathbf{x}'\boldsymbol{\beta})$.
  - But make sure standard errors etc. are robust to $\mathsf{V}[y|\mathbf{x}] \neq \mathsf{E}[y|\mathbf{x}]$.
  - 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.

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

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 $\boldsymbol{\beta}$.

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)
  - this robustness holds for only some likelihood based models.

Poisson regression: distribution of Poisson MLE

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

$$\widehat{\mathrm{V}}_{\mathrm{MLE}}[\widehat{\boldsymbol{\beta}}] = \left(\sum_i \widehat{\mu}_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \quad \text{using } \partial^2 \ln L(\boldsymbol{\beta}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}' = -\sum_i \exp(\mathbf{x}_i' \boldsymbol{\beta}))$$

- 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 $\widehat{\boldsymbol{\beta}} \overset{a}{\sim} \mathcal{N}[\boldsymbol{\beta}, \mathrm{V}_{\mathrm{ROB}}[\widehat{\boldsymbol{\beta}}]]$ 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 $\alpha >> 1$.

ASIDE: derivation of robust sandwich

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

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

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

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

where $\mu_i = \exp(\mathbf{x}_i'\boldsymbol{\beta})$, $\widehat{\mu}_i = \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})$ 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 $\sigma_i^2 = \mu_i$.

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

Poisson regression: coefficient interpretation

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.
- 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

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

- 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
  - 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
- Old Stata 10: add-ons `mfx` for MEM and `margeff` for AME.

- 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)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   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
------------------------------------------------------------------------------
```

Factor variables

- Problem: obtaining ME in models with interactions and polynomials
- Solution: Factor variable operators # and ##
  - also i. for discrete variables uses finite difference not derivatives
- Now can get ME with respect to age allowing for age^2 in model.
- Also now the i. variables have a somewhat different ME.

```
. * 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. Summary of remaining topics

- The Poisson is a **generalized linear model** (GLM)
  - the framework used in the statistics literature for nonlinear regression
  - leading examples are OLS, logit, probit, gamma regression
  - in Stata use command `glm` with family member `poisson` and `log` ling function.
- The Poisson MLE is generally inefficient
  - but generally not great efficiency loss (like OLS versus GLS)
- But it is usually the wrong model if we want to predict probabilities
  - intuitively Poisson has only one parameter μ whereas e.g. the normal has μ and σ^2
- The standard better ML model is the **negative binomial distribution**
  -
    - then $Var(y) = \mu + \alpha\mu^2$ (overdispersion) versus Poisson $Var(y) = \mu$.
    - in Stata use command `nbreg`.

4. Summary of remaining topics (continued)

- Count data may be record no zeroes
  - e.g. record number of doctor visits only for those who visited the clinic at least once
  - then use truncated MLE.
- Count data may be topcoded
  - e.g. record number of doctor visits as 0, 1, 2, 3, 4 0r more.
  - then use censored (from above) MLE.
- Count data may be interval recorded.
- Count data even if fully observed often have "too few zeroes"
  - e.g. fewer zeroes than predicted by a negative binomial model.
  - then two approaches
    - ★ hurdle model - one model for $= 0$ or > 0 and separate model given > 0
    - ★ with zeroes model
- Most natural extension of R^2 is $R^2_{\text{Cor}} = \widehat{\text{Cor}}^2[y_i, \widehat{y}_i]$.
- Compare nonnested parametric models using AIC or BIC.
- Quantile regression has been extended to counts.

5. Generalized linear models

- Generalized linear models (GLM) is the framework used in the statistics literature for nonlinear regression
  - a brief introduction that may be skipped depending on time.
- 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 robust standard errors: `vce(robust)`

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
```

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 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

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

- 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 s.e.'s or
  - Default ML s.e.'s multiplied by $\sqrt{\widehat{\alpha}}$ where $\mathsf{V}[y_i|\mathbf{x}_i] = \alpha \times h(\mathsf{E}[y_i|\mathbf{x}_i])$.

ASIDE: Nonlinear least squares estimator

- Alternative estimator for counts that is not used, because Poisson is simpler and is usually more efficient.
- Specify same conditional mean as Poisson: $\mathrm{E}[y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta})$.
- Minimize sum of squared residuals: $\sum_{i=1}^{N}(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))^2$.
- NLS is consistent provided $\mathrm{E}[y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta})$
- $\widehat{\boldsymbol{\beta}}_{\mathrm{NLS}} \overset{a}{\sim} \mathcal{N}[\boldsymbol{\beta}, \mathrm{V}_{\mathrm{MLE}}[\widehat{\boldsymbol{\beta}}]]$ and use robust sandwich variance estimate:

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

- For doctor visits data
  - `nl (docvis = exp({xb:  $xlist}+{b0})), vce(robust)`
  - NLS robust standard errors are 5-20% larger than those for Poisson

6. Diagnostics: residuals and influence measures

- Some diagnostics come out of the GLM literature.
- 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: $\mathbf{H} = \mathbf{W}^{1/2}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{1/2}$; $\mathbf{W} = \mathrm{Diag}[\widehat{\mu}_i]$.
  - 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 $\widehat{\boldsymbol{\beta}}$ 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

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

- 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

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

ASIDE: Diagnostics: overdispersion test

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

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

  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 tests (such as LM of Poisson vs. negative binomial) only test overdispersion.
- If we are just modelling the conditional mean, overdispersion is okay provided robust standard errors are calculated.

ASIDE:

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

- Now suppose we want to predict probability of 0 doctor visits, 1 doctor visits, ...
- 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 $\bar{p}_j$, $j = 0, 1, 2, ...$
- 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 $NR_u^2 > \chi^2_\alpha(J-1)$ where J is number of cells.
- Use Stata add-on chi2gof: e.g. `chi2gof, cells(11) table`

Stata add-on `chi2gof, cells(11) table` yields $\chi^2_\alpha(10)$ statistic equals 1103.43.

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

```
. chi2gof, cells(11) table

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

         Chi-square chi2(10) = 1103.43
         Prob>chi2           =    0.00
```

Cells	Abs. Freq.	Rel. Freq.	Fitted Rel. Freq.	Abs. Dif.
0	401	.1091	.0074	.1017
1	314	.0854	.0296	.0558
2	358	.0974	.063	.0344
3	334	.0908	.0954	.0045
4	339	.0922	.1159	.0237
5	266	.0723	.1212	.0489
6	231	.0628	.114	.0511
7	202	.0549	.0994	.0444
8	179	.0487	.0821	.0335
9	154	.0419	.0655	.0236
10 or more	791	.2445	.2067	.0378

(supersedes addon `countfit`: compares actual and fitted frequencies but no test).

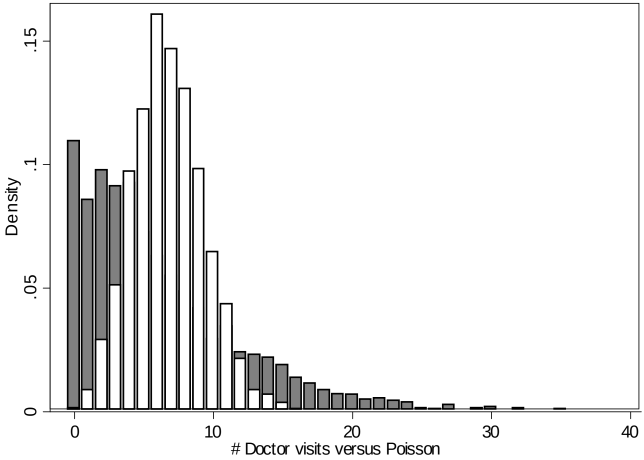
7. Negative binomial regression: motivation

- 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.

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

  then

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

- Probability mass function:

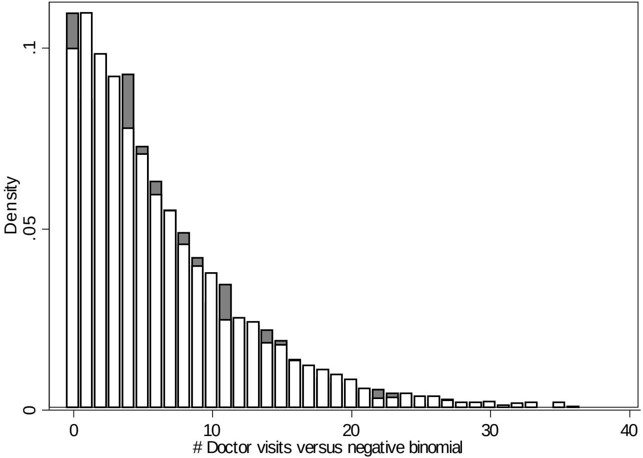
$$\Pr[Y = y|\lambda, \alpha] = \frac{\Gamma(\alpha^{-1} + y)}{\Gamma(\alpha^{-1})\Gamma(y + 1)} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda}\right)^{\alpha^{-1}} \left(\frac{\lambda}{\lambda + \alpha^{-1}}\right)^{y}.$$

- Mean and variance:

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

- 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 (from output not given)
`chi2gof, cells(11) table` yields $\chi^2_\alpha(10)$ statistic equals 53.62.

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

8. Richer parametric models

- Data frequently exhibit "non-Poisson" features:
  - Overdispersion: conditional variance exceeds conditional mean whereas Poisson imposes equality.
  - Excess zeros: higher frequency of zeros than predicted by Poisson.
- This provides motivation for richer parametric models than basic Poisson.
- Some models still have $\mathrm{E}[y|\mathbf{x}] = \exp(\mathbf{x}'\boldsymbol{\beta})$
  - Then richer model can provide more efficient estimates.
- Other models imply $\mathrm{E}[y|\mathbf{x}] \neq \exp(\mathbf{x}'\boldsymbol{\beta})$
  - Then Poisson QMLE is inconsistent
  - And marginal effects and coefficient interpretation more difficult.
- Many of these models are fully parametric and require correct specification for consistency.

Counts left-truncated at zero

- Sampling rule is such that observe only y and $\mathbf{x}$ for $y \geq 1$ i.e. only those who participate at least once are in sample.
- Truncated density (given untruncated density $f(y|\mathbf{x},\boldsymbol{\theta})$) is

$$f(y|\mathbf{x},\boldsymbol{\theta},y\geq 1)=\frac{f(y|\mathbf{x},\boldsymbol{\theta})}{\Pr[y\geq 0|\mathbf{x},\boldsymbol{\theta}]}=\frac{f(y|\mathbf{x},\boldsymbol{\theta})}{[1-f(0|\mathbf{x},\boldsymbol{\theta})]}.$$

- MLE is inconsistent if any aspect of the parametric model is misspecified.
- Need to assume that the process for nonzeroes is the same as zeroes.
  - e.g. If data are on annual number of hunting trips for only those who hunted this year, then a missing 0 is interpreted as being for a hunter who did not hunt this year (rather than for all people).
- Stata commands `ztp` (poisson) and `ztnb` (negative binomial)
  - `ztnb docvis $xlist if docvis>0, nolog`

Counts right-censored

- Sampling rule is that observe only 0, 1, 2, ..., $c-1$, c or more i.e. Only record counts up to c and then any value above c.
- Censored density (given uncensored density $f(y|\mathbf{x},\boldsymbol{\theta})$ and cdf is $F(y|\mathbf{x},\boldsymbol{\theta})$)

$$\begin{cases} f(y|\mathbf{x},\boldsymbol{\theta}) & y \leq c-1 \\ 1-F(c-1|\mathbf{x},\boldsymbol{\theta}) = 1-\sum_{j=0}^{c-1} f(j|\mathbf{x},\boldsymbol{\theta}) & y = c \end{cases}$$

- Log-likelihood (where $d_i = 1$ if uncensored and $d_i = 0$ if censored)

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \{d_i \ln f(y_i|\mathbf{x}_i,\boldsymbol{\theta}) + (1-d_i)\ln(1-\sum_{j=0}^{c-1} f(j|\mathbf{x}_i,\boldsymbol{\theta}))\}$$

- MLE is inconsistent if any aspect of the parametric model is misspecified
  - So pick a good density - at least negative binomial.
- Stata code up yourself
  - using command `ml` (for user-defined likelihood).

Counts recorded in intervals

- Sampling rule is that observe only counts in ranges.
  e.g. 0, 1-4, 5-9, 10 and above.
- Interval density is simply

$$\Pr[a \leq y \leq b] = \sum_{j=a}^{b} f(j|\mathbf{x}, \boldsymbol{\theta}).$$

- Let interval ranges by $[a_0, a_1 - 1]$, $[a_1, a_2 - 1]$, ..., $[a_m, a_{m+1})$, where $a_0 = 0$, $a_{m+1} = \infty$.
  Let d_k be binary indicators for whether in interval k $(k = 0, ..., m)$.
  Then

$$\ln L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \left[\sum_{k=0}^{m} d_{ij} \ln \left(\sum_{k=a_k}^{a_{k+1}-1} f(j|\mathbf{x}, \boldsymbol{\theta}) \right) \right].$$

- MLE is inconsistent if any aspect of the parametric model is misspecified.
- Stata has no command so need to code up.
- For convenience could instead use ordered logit or probit here.

Hurdle model or two-part model

- Suppose zero counts are determined by a different process to positive counts.
  - Zeros: density $f_1(y|\mathbf{x}_1, \boldsymbol{\theta}_1)$ so $\Pr[y=0] = f_1(0)$ and $\Pr[y>0] = 1 - f_1(0)$.
  - Positives: density $f_2(y|\mathbf{x}_2, \boldsymbol{\theta}_2)$ so truncated density $f_2(y)/(1-f_2(0))$.
- e.g. First - do I hunt this year or not?
  Second - given I chose to hunt, how many times (≥ 1)?
- Combined density is

$$f(y|\mathbf{x}_1, \mathbf{x}_1, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \begin{cases} f_1(y|\mathbf{x}_1, \boldsymbol{\theta}_1) & y = 0 \\ \dfrac{1 - f_1(0|\mathbf{x}_1, \boldsymbol{\theta}_1)}{1 - f_2(0|\mathbf{x}_2, \boldsymbol{\theta}_2)} \times f_2(y|\mathbf{x}_2, \boldsymbol{\theta}_2) & y \geq 1 \end{cases}$$

- MLE is inconsistent if any aspect of model misspecified.

- Conditional mean is now

$$\mathrm{E}[y|\mathbf{x}] = \Pr[y_1 > 0|\mathbf{x}_1] \times \mathrm{E}_{y_2>0}[y_2|y_2 > 0, \mathbf{x}_2].$$

- This makes marginal effects more complicated.
- Example: $f_1(\cdot)$ is logit and $f_2(\cdot)$ is negative binomial.
- Then

$$\mathrm{E}[y|\mathbf{x}] = \Lambda(\mathbf{x}_1'\boldsymbol{\beta}) \times \exp(\mathbf{x}_2'\boldsymbol{\beta})/[1 - (1 + \alpha_2 \exp(\mathbf{x}_2'\boldsymbol{\beta}))^{-1/\alpha_2}],$$

where $\Lambda(z) = e^z/(1 + e^z)$.

Hurdle model - logit and negative binomial: Stata addon hnblogit

```
. hnblogit docvis $xlist, nolog

Negative Binomial-Logit Hurdle Regression          Number of obs   =       3677
                                                   Wald chi2(7)    =     309.90
Log likelihood = -10493.225                        Prob > chi2     =     0.0000
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
logit						
private	.6586978	.1264608	5.21	0.000	.4108393	.9065563
medicaid	.0554225	.1726694	0.32	0.748	-.2830032	.3938483
age	.542878	.2238845	2.42	0.015	.1040724	.9816835
age2	-.0034989	.0014957	-2.34	0.019	-.0064304	-.0005673
educyr	.047035	.0155706	3.02	0.003	.0165171	.0775529
actlim	.1623927	.1523743	1.07	0.287	-.1362554	.4610408
totchr	1.050562	.0671922	15.64	0.000	.9188676	1.182256
_cons	-20.94163	8.335138	-2.51	0.012	-37.2782	-4.605058
negbinomial						
private	.1095566	.0345239	3.17	0.002	.041891	.1772222
medicaid	.0972308	.0470358	2.07	0.039	.0050423	.1894193
age	.2719031	.0625359	4.35	0.000	.149335	.3944712
age2	-.0017959	.000416	-4.32	0.000	-.0026113	-.0009805
educyr	.0265974	.0043937	6.05	0.000	.0179859	.035209
actlim	.1955384	.0355161	5.51	0.000	.125928	.2651487
totchr	.2226967	.0124128	17.94	0.000	.1983681	.2470252
_cons	-9.190165	2.337592	-3.93	0.000	-13.77176	-4.608569
/lnalpha	-.525962	.0418671	-12.56	0.000	-.60802	-.443904

```
AIC Statistic =       5.712
```

Zero-inflated model (or with-zeroes model)

- Suppose there is an additional reason for zero counts
  - Extra model for 0: density $f_1(y|\mathbf{x}_1, \boldsymbol{\theta}_1)$
  - Usual model for 0: realization of 0 from density $f_2(y|\mathbf{x}_2, \boldsymbol{\theta}_2)$.
- e.g. Some zeroes are mismeasurement and some are true zeros.
- Zero-inflated model has density

$$
\begin{aligned}
& f(y|\mathbf{x}_1, \mathbf{x}_1, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\
= & \begin{cases} f_1(0|\mathbf{x}_1, \boldsymbol{\theta}_1) + [1 - f_1(0|\mathbf{x}_1, \boldsymbol{\theta}_1)] \times f_2(0|\mathbf{x}_2, \boldsymbol{\theta}_2) & y = 0 \\ [1 - f_1(0|\mathbf{x}_1, \boldsymbol{\theta}_1)] \times \cdot f_2(y|\mathbf{x}_2, \boldsymbol{\theta}_2) & y \geq 1 \end{cases}
\end{aligned}
$$

- MLE is inconsistent if any aspect of model misspecified.
- Not used much in econometrics - hurdle model more popular.

Zero-inflated negative binomial: Stata command zinb and zip

```
. zinb docvis $xlist, inflate($xlist) vuong nolog

Zero-inflated negative binomial regression        Number of obs     =      3677
                                                  Nonzero obs       =      3276
                                                  Zero obs          =       401

Inflation model = logit                           LR chi2(7)        =    588.93
Log likelihood  = -10492.88                       Prob > chi2       =    0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
docvis       |
     private |   .1289797    .032987     3.91   0.000     .0643264     .193633
    medicaid |   .1091956    .044511     2.45   0.014     .0219556    .1964356
         age |   .2847325   .0589577     4.83   0.000     .1691776    .4002874
        age2 |  -.0018781   .0003922    -4.79   0.000    -.0026469   -.0011093
      educyr |   .0253991   .0041432     6.13   0.000     .0172786    .0335196
      actlim |   .1737716   .0336464     5.16   0.000     .1078258    .2397173
      totchr |    .229991   .0120795    19.04   0.000     .2063156    .2536663
       _cons |  -9.680235   2.204161    -4.39   0.000    -14.00031    -5.36016
-------------+----------------------------------------------------------------
inflate      |
     private |  -.9152675   .2758402    -3.32   0.001    -1.455904   -.3746307
    medicaid |   .3487142   .3372848     1.03   0.301    -.3123519     1.00978
         age |  -.4357439   .5156094    -0.85   0.398     -1.44632    .5748319
        age2 |    .002805   .0034886     0.80   0.421    -.0040326    .0096426
      educyr |    -.08423   .0339273    -2.48   0.013    -.1507263   -.0177336
      actlim |  -.8241735   .4825621    -1.71   0.088    -1.769978    .1216309
      totchr |  -2.985208   .6860952    -4.35   0.000     -4.32993   -1.640486
       _cons |   17.09618   18.97318     0.90   0.368    -20.09057    54.28294
-------------+----------------------------------------------------------------
    /lnalpha |  -.5848279   .0349792   -16.72   0.000    -.6533859   -.5162699
-------------+----------------------------------------------------------------
       alpha |   .5572017   .0194905                      .5202812    .5967423
------------------------------------------------------------------------------
Vuong test of zinb vs. standard negative binomial: z =     6.48  Pr>z = 0.0000
```

Continuous mixture models

- Mixture motivation for negative binomial assumes $y|\theta \sim Poisson\ (\theta)$ where $\theta = \lambda \upsilon$ is the product of two components:
  - observed individual heterogeneity $\lambda = \exp(\mathbf{x}'\boldsymbol{\beta})$
  - unobserved individual heterogeneity $\upsilon \sim Gamma[1,\ \alpha]$.
- Integrating out

$$h(y|\lambda) = \int f(y|\lambda, \upsilon)g(\upsilon)d\upsilon = \int [e^{-\lambda\upsilon}(\lambda\upsilon)^y / y!] \times g(\upsilon)d\upsilon$$

  gives $y|\lambda \sim NB\ [\lambda,\ \lambda + \alpha\lambda^2]$ if $\upsilon \sim Gamma[1,\ \alpha]$.
- Different distributions of υ lead to different models
  - e.g. Poisson-lognormal mixture (random effects model)
  - e.g. Poisson-Inverse Gaussian.
- Even if no closed form solution can estimate using
  - numerical integration (one-dimensional) e.g. Gaussian quadrature.
  - Monte Carlo integration e.g. maximum simulated likelihood.

Hierarchical models

- For multi-level surveys cross-section data individuals i may be in cluster j
  - e.g. patient i in hospital j
  - e.g. individual i in household j or village j
- Hierarchical model or generalized linear mixed model example

$$
\begin{aligned}
y_i &\sim \text{Poisson}[\mu_{ij} = \exp(\mathbf{x}'_{ij}\boldsymbol{\beta}_j + \varepsilon_{ij})] \\
\boldsymbol{\beta}_j &= \mathbf{W}_j\gamma + \mathbf{v}_j \\
\varepsilon_{ij} &\sim \mathcal{N}[0, \sigma^2_\varepsilon] \\
\mathbf{v}_j &\sim \mathcal{N}[\mathbf{0}, \text{Diag}[\sigma^2_{jk}]]
\end{aligned}
$$

  - Estimate by MLE or by Bayesian methods
  - Stata command `mepoisson`

Model comparison for fully parametric models

- Choice between nested models using likelihood ratio tests
  - e.g. Poisson versus negative binomial.
- Choice between non-nested models using Vuong's (1989) likelihood ratio test
  - e.g. Zero-inflated NB versus NB
- Choice between non-nested mixture models using penalized log-likelihood
  - Akaike's information criterion (AIC) and extensions ($q = \#$ parameters)

$$\begin{aligned} AIC &= -2\ln L + 2q \\ BIC &= -2\ln L + qk\ln N \\ CAIC &= -2\ln L + q(1+\ln)N \end{aligned}$$

  - Prefer model with small AIC or BIC.
  - AIC penalty for larger model too small. Bayesian IC (BIC) better.

- Compare predicted means: $\mathrm{E}[y|\mathbf{x}, \widehat{\boldsymbol{\theta}}]$.
- Compare observed frequencies $\overline{p}_j$ to average predicted frequencies

$$\widehat{p}_j = N^{-1} \sum_{i=1}^{N} \widehat{p}_{ij},$$

where $\widehat{p}_{ij} = \widehat{\Pr}[y_i = j]$.

Compare AIC, BIC for regular NB, hurdle logit/NB and zero-inflated NB.

Statistics	NBREG	HURDLENB	ZINB
N	3677	3677	3677
ll	-10589.3	-10493.2	-10492.9
aic	21196.7	21020.4	21019.8
bic	21252.6	21126.0	21125.3

Hurdle NB and ZINB are big improvement on regular NB
- lnL is approximately 100 higher than for NB
- AIC and BIC is much smaller (with only 9 extra parameters)
Little difference between Hurdle NB and ZINB.

The conditional means from the three models are similar.

```
. summarize docvis dvnbreg dvhurdle dvzinb
```

Variable	Obs	Mean	Std. Dev.	Min	Max
docvis	3677	6.822682	7.394937	0	144
dvnbreg	3677	6.890034	3.486562	2.078925	41.31503
dvhurdle	3677	6.840676	3.134925	1.35431	31.86874
dvzinb	3677	6.838704	3.135122	.9473827	32.98153

```
. correlate docvis dvnbreg dvhurdle dvzinb
(obs=3677)
```

	docvis	dvnbreg	dvhurdle	dvzinb
docvis	1.0000			
dvnbreg	0.3870	1.0000		
dvhurdle	0.3990	0.9894	1.0000	
dvzinb	0.3983	0.9882	0.9982	1.0000

Quantile regression

- The q^{th} quantile regression estimator $\widehat{\boldsymbol{\beta}}_q$ minimizes over $\boldsymbol{\beta}_q$

$$Q(\boldsymbol{\beta}_q) = \sum_{i:y_i \geq \mathbf{x}_i'\boldsymbol{\beta}}^{N} q|y_i - \mathbf{x}_i'\boldsymbol{\beta}_q| + \sum_{i:y_i < \mathbf{x}_i'\boldsymbol{\beta}}^{N} (1-q)|y_i - \mathbf{x}_i'\boldsymbol{\beta}_q|, \quad 0 < q < 1.$$

  - Example: median regression with $q = 0.5$.
- For count y adapt standard methods for continuous y by:
  - Replace count y by continuous variable $z = y + u$ where $u \sim Uniform[0, 1]$.
  - Then reconvert predicted z-quantile to y-quantile using ceiling function.
  - Machado and Santos Silva (2005).
- Stata example
  - `qcount docvis $xlist, q(0.5) rep(50)`
  - `qcount_mfx // MEM`

9. 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.

10. References: General sources that include counts

- 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 (2010)
  "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.

More Specific 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.
  - 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 (2012), *Generalized Linear Models and Extensions*, Third Edition, Stata Press.