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)
  - \( x \): 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 \( E[y|x] \):
  ▶ 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:
  ▶ \( E[y|x] = \exp(x'\beta) \) so \( ME_j = \frac{\partial E[y|x]}{\partial x_j} = \beta_j \exp(x'\beta) \neq \beta_j \).

• 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.
1. Introduction

2. Poisson (with cross-section data) Theory*

3. Poisson Application*

4. Summary of the remaining topics*

5. Generalized linear models

6. Diagnostics

7. Negative binomial model

8. Richer Parametric Model (censored, truncated, hurdle, ..)

9. Summary

10. References
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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>docvis</td>
<td>3677</td>
<td>6.822682</td>
<td>7.394937</td>
<td>0</td>
<td>144</td>
</tr>
</tbody>
</table>
```

A. Colin Cameron, Univ. of Calif. - Davis ... for Center of Labor Economics, Norwegian School of Economics. Advanced Microeconometrics ()
Count Regression: Part 1 Aug 28 - Sept 1, 2017 5 / 65
- Doctor visits: Histogram dropping observations with more than 40 visits
Poisson distribution

- From stochastic process theory, natural model for counts is
  \[ y \sim \text{Poisson}[\lambda]. \]

- Probability mass function:
  \[ \Pr(Y = y|\lambda) = \frac{e^{-\lambda} \lambda^y}{y!} \]

- Mean and variance:
  \[ E[y] = \lambda \quad \text{and} \quad V[y] = \lambda \]

- Equidispersion: variance = mean
  - Restriction imposed by Poisson

- Overdispersion: variance > mean
  - More common feature of count data
  - Doctor visits data: \( \bar{y} = 6.82, \ s_y^2 = 54.68 \approx 8.01\bar{y}. \)
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 $E[y|x] = \exp(x'\beta)$.
  - But make sure standard errors etc. are robust to $V[y|x] \neq E[y|x]$.
  - And generally don’t use Poisson if need to predict probabilities.
Poisson regression: Poisson MLE

- Let the Poisson rate parameter vary across individuals with \( x \) in way to ensure \( \lambda > 0 \).
  \[
  \lambda = E[y|x] = \exp(x' \beta).
  \]

- MLE is straightforward given data independent over \( i \).
  \[
  f(y) = e^{-\lambda} \frac{\lambda^y}{y!}
  \]
  \[
  \Rightarrow \ln f(y) = -\exp(x' \beta) + yx' \beta - \ln y!
  \]
  \[
  \Rightarrow \ln L(\beta) = \sum_{i=1}^{n} \{ -\exp(x_i' \beta) + y_i x_i' \beta - \ln y_i! \}
  \]
  \[
  \Rightarrow \frac{\partial \ln L(\beta)}{\partial \beta} = \sum_{i=1}^{n} \{ -\exp(x_i' \beta)x_i + y_i x_i \} \]
The ML first-order conditions are

\[ \sum_{i=1}^{n} (y_i - \exp(x_i \hat{\beta})) x_i = 0. \]

No explicit solution for \( \hat{\beta} \).

- Instead use Newton-Raphson iterative method.
- Fast as objective function is globally concave in \( \beta \).
Poisson regression: consistency of Poisson MLE

- ML first-order conditions are
  \[ \sum_{i=1}^{n} (y_i - \exp(x_i' \hat{\beta})) x_i = 0. \]

- Consistency only requires (given independence over \(i\))
  \[ \mathbb{E}[(y_i - \exp(x_i' \beta)) x_i] = 0 \]

- So consistent if
  \[ \mathbb{E}[y_i | x_i] = \exp(x_i' \beta) \]

- 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 $\hat{\beta} \sim \mathcal{N}[\beta, V_{\text{MLE}}[\beta]]$ where

$$
\hat{V}_{\text{MLE}}[\beta] = \left( \sum_i \hat{\mu}_i x_i x_i' \right)^{-1}
$$

using $\partial^2 \ln L(\beta)/\partial \beta \partial \beta' = - \sum_i \exp(x_i' \beta))$

- If distribution is not Poisson but $E[y_i | x_i] = \exp(x_i' \beta)$ and $V[y_i | x_i] = \sigma_i^2$ then $\hat{\beta} \sim \mathcal{N}[\beta, V_{\text{ROB}}[\beta]]$ and we use the robust sandwich estimate of variance (White (1982), Huber (1967))

$$
\hat{V}_{\text{ROB}}[\beta] = \left( \sum_i \hat{\mu}_i x_i x_i' \right)^{-1} \left( \sum_i (y_i - \hat{\mu}_i)^2 x_i x_i' \right) \left( \sum_i \hat{\mu}_i x_i x_i' \right)^{-1}
$$

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

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

Count Regression: Part 1  
Aug 28 - Sept 1, 2017  12 / 65
ASIDE: derivation of robust sandwich

- Take a first-order Taylor series expansion of \( \sum_i (y_i - \exp(x_i'\hat{\beta}))x_i \) about \( \beta \).

\[
\sum_i (y_i - \exp(x_i'\hat{\beta}))x_i = \sum_i (y_i - \exp(x_i'\beta))x_i - \sum_i \exp(x_i'\beta)x_i x_i'(\hat{\beta} - \beta) - R.
\]

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

\[
\Rightarrow \sum_i (y_i - \mu_i)x_i + (\sum_i - \mu_i x_i x_i') (\hat{\beta} - \beta) = 0
\]

\[
(\hat{\beta} - \beta) \sim (\sum_i \mu_i x_i x_i')^{-1} \times \sum_i (y_i - \mu_i)x_i
\]

\[
\sim \mathcal{N} \left[ 0, (\sum_i \mu_i x_i x_i')^{-1} \left( \sum_i \sigma_i^2 x_i x_i' \right) (\sum_i \mu_i x_i x_i')^{-1} \right]
\]

where \( \mu_i = \exp(x_i'\beta) \), \( \hat{\mu}_i = \exp(x_i'\hat{\beta}) \) and \( \sigma_i^2 = \text{E}[(y_i - \mu_i)^2] \).

- Asymptotically can estimate \( \sum_i \sigma_i^2 x_i x_i' \) by \( (\sum_i (y_i - \hat{\mu}_i)^2 x_i x_i') \).

- If density is Poisson then simplifies to \( (\sum_i \mu_i x_i x_i')^{-1} \) 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

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

: summarize docvis $xlist, sep(10)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>docvis</td>
<td>3677</td>
<td>6.822682</td>
<td>7.394937</td>
<td>0</td>
<td>144</td>
</tr>
<tr>
<td>private</td>
<td>3677</td>
<td>0.4966005</td>
<td>0.5000564</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>medicaid</td>
<td>3677</td>
<td>0.166712</td>
<td>0.3727692</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>age</td>
<td>3677</td>
<td>74.24476</td>
<td>6.376638</td>
<td>65</td>
<td>90</td>
</tr>
<tr>
<td>age2</td>
<td>3677</td>
<td>5552.936</td>
<td>958.9996</td>
<td>4225</td>
<td>8100</td>
</tr>
<tr>
<td>educyr</td>
<td>3677</td>
<td>11.18031</td>
<td>3.827676</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>actlim</td>
<td>3677</td>
<td>0.333152</td>
<td>0.4714045</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>totchr</td>
<td>3677</td>
<td>1.843351</td>
<td>1.350026</td>
<td>0</td>
<td>8</td>
</tr>
</tbody>
</table>
```
Poisson MLE with robust sandwich standard errors - preferred

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

Poisson regression

| Coef.   | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|---------|-----------|-------|------|---------------------|
| _cons  | -10.18221 | 2.369212 | -4.30 | 0.000 | -14.82578 | -5.538638 |
| totchr | 0.2483898 | 0.0125786 | 19.75 | 0.000 | 0.2237361 | 0.2730435 |
| actlim | 0.1864213 | 0.0396569 | 4.70 | 0.000 | 0.1086953 | 0.2641474 |
| educyr | 0.0295562 | 0.0048454 | 6.10 | 0.000 | 0.0200594 | 0.0390530 |
| age2   | -0.0019311 | 0.0004166 | -4.64 | 0.000 | -0.0027475 | -0.0011147 |
| age    | 0.2936722 | 0.0629776 | 4.66 | 0.000 | 0.1702383 | 0.4171061 |
| medicaid | 0.0970005 | 0.0568264 | 1.71 | 0.088 | -0.012773 | 0.2083783 |
| private | 0.1422324 | 0.036356 | 3.91 | 0.000 | 0.070976 | 0.2134889 |
| private | 0.0970005 | 0.0568264 | 1.71 | 0.088 | -0.0143773 | 0.2083783 |
| age2   | -0.0019311 | 0.0004166 | -4.64 | 0.000 | -0.0027475 | -0.0011147 |
| educyr | 0.0295562 | 0.0048454 | 6.10 | 0.000 | 0.0200594 | 0.0390530 |
| age    | 0.2936722 | 0.0629776 | 4.66 | 0.000 | 0.1702383 | 0.4171061 |
| medicaid | 0.0970005 | 0.0568264 | 1.71 | 0.088 | -0.012773 | 0.2083783 |
| private | 0.1422324 | 0.036356 | 3.91 | 0.000 | 0.070976 | 0.2134889 |

Log pseudolikelihood = -15019.64

Poisson regression

Number of obs = 3677
Wald chi2(7) = 720.43
Prob > chi2 = 0.0000
Pseudo R2 = 0.1297
3. Poisson regression example

Poisson MLE with default ML standard errors - do not use
- These are misleadingly small due to overdispersion!!

. * Poisson with default ML standard errors
.poison 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
Log likelihood =  -15019.64                      Prob > chi2   =  0.0000
Number of obs   = 3677                            LR chi2( 7)   =  4477.98
Iteration 2:    log likelihood =  -15019.64 
Iteration 1:    log likelihood =  -15019.64 
Iteration 0:    log likelihood = -15019.656  

|            | Coef.  | Std. Err. |    z  |   P>|z|  |      [95% Conf. Interval]       |
|------------|--------|-----------|------|-------|---------------------------------|
| _cons      | -10.18221 | .9720115 | -10.48 | 0.000 | -12.08732 to -8.277101          |
| totchr     | .2483898 | .0046447 | 53.48 | 0.000 | .2392864 to .2574933            |
| actlim     | .1864213 | .014566  | 12.80 | 0.000 | .1578726 to .2149701           |
| educyr     | .2483898 | .0046447 | 53.48 | 0.000 | .2392864 to .2574933            |
| age2       | -.0019311 | .001724  | -11.20 | 0.000 | -.0022691 to -.0015931         |
| age        | .2936722 | .0259563 | 11.31 | 0.000 | .2427988 to .3445457            |
| medicaid   | .0970005 | .0189307 | 5.12  | 0.000 | .0598969 to .134104             |
| private    | .1422324 | .0143311 | 9.92  | 0.000 | .114144 to .1703208             |
| private    | .0970005 | .0189307 | 5.12  | 0.000 | .0598969 to .134104             |
| age        | .2936722 | .0259563 | 11.31 | 0.000 | .2427988 to .3445457            |
| medicaid   | .0970005 | .0189307 | 5.12  | 0.000 | .0598969 to .134104             |
| private    | .1422324 | .0143311 | 9.92  | 0.000 | .114144 to .1703208             |

Robust se’s are 2.5-2.7 times larger

Note: \[ \sqrt{\frac{s_y^2}{\bar{y}}} = \sqrt{7.39^2/6.82} = \sqrt{8.01} = 2.830. \]
For the exponential conditional mean the marginal effect

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

1. Conditional mean is strictly monotonic increasing (or decreasing) in $x_{ij}$ according to the sign of $\beta_j$.

2. Coefficients are semi-elasticities:
   $\beta_j$ 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 $x_k$. 
Example: $\hat{\beta}_{\text{Private}} = 0.142$.

- 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: $\hat{\beta}_{\text{Private}} = 0.142$ and $\hat{\beta}_{\text{totchr}} = 0.248$

- Since $0.142/0.248 = 0.57$, private insurance has the same impact on mean doctor visits as 0.57 more chronic conditions.
3. Poisson regression example

Marginal effects: Three types

1. Average marginal effect (AME): Evaluate at each \( x_i \) and average

\[
AME = \sum_i \frac{\partial E[y_i|x_i]}{\partial x_{ij}} = \sum_i \exp(x_i' \hat{\beta}) \times \hat{\beta}_j.
\]

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

\[
MEM = \frac{\partial E[y|x]}{\partial x_j} \bigg|_{x=\bar{x}} = \exp(\bar{x}' \hat{\beta}) \times \hat{\beta}_j
\]

3. Marginal effect at representative value (MER): Evaluate at \( x = x^* \)

AME is nest

- For population AME use population weights in computing AME.
For Poisson with intercept in model $\text{AME} = \bar{y} \hat{\beta}_j$

- Reason: f.o.c. $\sum_i (y_i - \exp(x_i'\hat{\beta})) = 0$ imply $\sum_i \exp(x_i'\hat{\beta}) = \bar{y}$
- For Poisson can show that $\text{AME} > \text{MEM}$.

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

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
</tr>
<tr>
<td>private</td>
<td>.9704067</td>
<td>.247564</td>
<td>3.92</td>
<td>0.000</td>
<td>.4851902</td>
</tr>
<tr>
<td>medicaid</td>
<td>.6618034</td>
<td>.3901692</td>
<td>1.70</td>
<td>0.090</td>
<td>-.1029143</td>
</tr>
<tr>
<td>age</td>
<td>2.003632</td>
<td>.4303207</td>
<td>4.66</td>
<td>0.000</td>
<td>1.160219</td>
</tr>
<tr>
<td>age2</td>
<td>-.0131753</td>
<td>.0028473</td>
<td>-4.63</td>
<td>0.000</td>
<td>-.0187559</td>
</tr>
<tr>
<td>educyr</td>
<td>.2016526</td>
<td>.0337805</td>
<td>5.97</td>
<td>0.000</td>
<td>.1354441</td>
</tr>
<tr>
<td>actlim</td>
<td>1.271893</td>
<td>.2749286</td>
<td>4.63</td>
<td>0.000</td>
<td>.7330432</td>
</tr>
<tr>
<td>totchr</td>
<td>1.694685</td>
<td>.0908884</td>
<td>18.65</td>
<td>0.000</td>
<td>1.516547</td>
</tr>
</tbody>
</table>
3. Poisson regression example

Marginal effects: MEM

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

### Conditional marginal effects

<table>
<thead>
<tr>
<th>Expression</th>
<th>Number of obs = 3,677</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model VCE</td>
<td>Robust</td>
</tr>
</tbody>
</table>

#### Expression

```
Expression : Predicted number of events, predict()
```

#### dy/dx w.r.t.

- `private` = 0.4966005 (mean)
- `medicaid` = 0.166712 (mean)
- `age` = 74.24476 (mean)
- `age2` = 5552.936 (mean)
- `educyr` = 11.18031 (mean)
- `actlim` = 0.333152 (mean)
- `totchr` = 1.843351 (mean)

#### Delta-method

|                | dy/dx | Std. Err. | z      | P>|z| | [95% Conf. Interval] |
|----------------|-------|-----------|--------|------|----------------------|
| `private`      | 0.8914309 | 0.2270816 | 3.93   | 0.000 | 0.4463591 - 1.336503 |
| `medicaid`     | 0.607943 | 0.3577377 | 1.70   | 0.089 | -0.09321 - 1.309096  |
| `age`          | 1.840568 | 0.3924682 | 4.69   | 0.000 | 1.071345 - 2.609792  |
| `age2`         | -0.012103 | 0.0025973 | -4.66  | 0.000 | -0.0171936 - 0.0070125 |
| `educyr`       | 0.1852413 | 0.0306709 | 6.04   | 0.000 | 0.1251275 - 0.2453551 |
| `actlim`       | 1.168381 | 0.2516143 | 4.64   | 0.000 | 0.6752264 - 1.661536  |
| `totchr`       | 1.556764 | 0.0760166 | 20.48  | 0.000 | 1.407775 - 1.705754   |

A. Colin Cameron  Univ. of Calif. - Davis  Count Regression: Part 1  Aug 28 - Sept 1, 2017  23 / 65
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.

Note: dy/dx for factor levels is the discrete change from the base level.
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 link` 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 $\mu$ whereas e.g. the normal has $\mu$ and $\sigma^2$

The standard better ML model is the negative binomial distribution
- then \( \text{Var}(y) = \mu + \alpha \mu^2 \) (overdispersion) versus Poisson \( \text{Var}(y) = \mu \).
- 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 or 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 $R^2_{Cor} = \hat{\text{Cor}}^2 [y_i, \hat{y}_i]$.

- Compare nonnested parametric models using AIC or BIC.

- Quantile regression has been extended to counts.
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 $y \in \{0, 1, 2, 3, \ldots\}$
- 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
Optimization : ML
Deviance = 18395.14033
Pearson = 23147.37781

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

Log pseudolikelihood = -15019.6398

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

A. Colin Cameron  Univ. of Calif. - Davis  Count Regression: Part 1  Aug 28 - Sept 1, 2017  28 / 65

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

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

- For regression specify a model of the mean
  - \( \mu_i = \mu_i(\beta) = \mu_i(x_i, \beta) \).

- Poisson is a member with
  - \( a(\mu) = -\mu; c(\mu) = \ln \mu \)
  - \( a'(\mu) = -1 \) and \( c'(\mu) = 1 / \mu \)
  - \( E[y] = -(-1)/(1/\mu) = \mu \) and \( V[y] = 1/(1/\mu) = \mu \).
Quasi-MLE maximizes the log-likelihood

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

The first-order conditions are

\[
\sum_i \{a'(\mu_i(\beta)) + c'(\mu_i(\beta))y_i\} \times \frac{\partial \mu_i(\beta)}{\partial \beta} = 0
\]

\[
\Rightarrow \sum_i c'(\mu_i(\beta)) \times \left\{y_i - a'(\mu_i(\beta)) / c'(\mu_i(\beta))\right\} \times \frac{\partial \mu_i(\beta)}{\partial \beta} = 0
\]

\[
\Rightarrow \sum_i \frac{1}{\sqrt{y_i}} \left\{y_i - \mu_i(\beta)\right\} \times \frac{\partial \mu_i(\beta)}{\partial \beta} = 0
\]

MLE based on LEF with \( \mu_i = g(x_i'\beta) \) shares the robustness properties of normal and Poisson MLE

- consistency requires correct specification of the mean (so \( E[\{y_i - \mu_i(\beta)\}] = 0 \)).

But correct standard errors should use a robust estimate of variance

- Robust sandwich s.e.'s or
- Default ML s.e.'s multiplied by \( \sqrt{\alpha} \) where \( V[y_i|x_i] = \alpha \times h(E[y_i|x_i]) \).
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: $E[y_i|x_i] = \exp(x_i'\beta)$.
- Minimize sum of squared residuals: $\sum_{i=1}^{N}(y_i - \exp(x_i'\beta))^2$.
- NLS is consistent provided $E[y_i|x_i] = \exp(x_i'\beta)$
- $\hat{\beta}_{NLS} \sim \mathcal{N}[^\beta, \text{V}_{MLE}[^\beta]]$ and use robust sandwich variance estimate:

$$
\hat{\text{V}}_{\text{ROB}}[\hat{^\beta}_{NLS}] = \left(\sum_i \hat{\mu}_i^2 x_i x_i'\right)^{-1} \left(\sum_i (y_i - \hat{\mu}_i)^2 \hat{\mu}_i^2 x_i x_i'\right) \left(\sum_i \hat{\mu}_i^2 x_i x_i'\right)^{-1}.
$$

- 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: \( r_i = (y_i - \hat{\mu}_i) \)
  - Pearson: \( p_i = (y_i - \hat{\mu}_i) / \sqrt{\hat{\mu}_i} \)
  - Deviance: \( d_i = \text{sign}(y_i - \hat{\mu}_i) \sqrt{2 \left\{ y_i \ln(y_i / \hat{\mu}_i) - (y_i - \hat{\mu}_i) \right\}} \)
  - Anscombe: \( a_i = 1.5(y_i^{2/3} - \mu_i^{2/3}) / \mu_i^{1/6} \)
  - Last three will be standardized if \( \text{Var}[y_i] = \mu_i \).

- Small-sample corrections (for Poisson)
  - Hat matrix: \( H = W^{1/2}X(X'WX)^{-1}X'W^{1/2} \); \( W = \text{Diag}[\hat{\mu}_i] \).
  - 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 \( \hat{\beta} \) when observation \( i \) is omitted.
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 $\hat{y}_i = \hat{\mu}_i$

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

- Sums of squares measures differ in nonlinear models

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

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

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

  ▶ Works for Poisson as $\ln L_{max}$ occurs when $\mu_i = y_i$.
  ▶ Unlike others $0 \leq R^2_{RG} < 1$ and $R^2_{RG}$ always increases as add regressors.

- Stata measure is only applicable to binary and multinomial models

$$R^2_{Pseudo} = 1 - \frac{\ln L_{fit}}{\ln L_0}.$$
ASIDE: Diagnostics: overdispersion test

- $H_0 : \text{Var}[y_i|x_i] = \text{E}[y_i|x_i]$ versus $H_1 : \text{Var}[y_i|x_i] = \text{E}[y_i|x_i] + \alpha(\text{E}[y_i|x_i])^2$.
- Test $H_0 : \alpha = 0$ against $H_1 : \alpha > 0$.
- Implement by auxiliary regression

$$\frac{(y_i - \hat{\mu}_i)^2 - y_i}{\hat{\mu}_i} = \alpha \hat{\mu}_i + \text{error}$$

and do $t$ test of whether the coefficient of $\hat{\mu}_i$ 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.
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 $\bar{p}_j$ (fraction of observations with $y_i = j$).
- Fitted frequency $\hat{p}_j = N^{-1} \sum_{i=1}^{N} \hat{p}_{ij}$
  - predicted probability $\hat{p}_{ij} = Pr[y_i = j] = e^{-\mu_i} \mu_i^j / j!$ for Poisson.
- Expect $\hat{p}_j$ close to $\bar{p}_j$, $j = 0, 1, 2, ....$
- Informal statistic is Pearson’s chi-square test
  \[
  \sum_j \frac{(np_j - n\hat{p}_j)^2}{n\hat{p}_j}
  \]
  but this is not $\chi^2$ distributed due to estimation to get $\hat{p}_j$. 
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

$$1 = s_i(y_i, x_i, \hat{\theta})' \gamma + \sum_j (d_{ij}(y_i) - \hat{p}_{ij})' \delta_j + \text{error}$$

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
- $\hat{p}_{ij}$ equals predicted probability for that cell
- $s_i(y_i, x_i, \theta) = \partial \ln f(y_i|x_i, \theta)/\partial \theta \quad ( = (y_i - \exp(x_i'\beta))$ for Poisson).

Reject at level $\alpha$ 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
6. Diagnostics

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
```

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>401</td>
<td>.1091</td>
<td>.0074</td>
<td></td>
<td>.1017</td>
</tr>
<tr>
<td>1</td>
<td>314</td>
<td>.0854</td>
<td>.0296</td>
<td></td>
<td>.0558</td>
</tr>
<tr>
<td>2</td>
<td>358</td>
<td>.0974</td>
<td>.063</td>
<td></td>
<td>.0344</td>
</tr>
<tr>
<td>3</td>
<td>334</td>
<td>.0908</td>
<td>.0954</td>
<td></td>
<td>.0045</td>
</tr>
<tr>
<td>4</td>
<td>339</td>
<td>.0922</td>
<td>.1159</td>
<td></td>
<td>.0237</td>
</tr>
<tr>
<td>5</td>
<td>266</td>
<td>.0723</td>
<td>.1212</td>
<td></td>
<td>.0489</td>
</tr>
<tr>
<td>6</td>
<td>231</td>
<td>.0628</td>
<td>.114</td>
<td></td>
<td>.0511</td>
</tr>
<tr>
<td>7</td>
<td>202</td>
<td>.0549</td>
<td>.0994</td>
<td></td>
<td>.0444</td>
</tr>
<tr>
<td>8</td>
<td>179</td>
<td>.0487</td>
<td>.0821</td>
<td></td>
<td>.0335</td>
</tr>
<tr>
<td>9</td>
<td>154</td>
<td>.0419</td>
<td>.0655</td>
<td></td>
<td>.0236</td>
</tr>
<tr>
<td>10 or more</td>
<td>791</td>
<td>.2445</td>
<td>.2067</td>
<td></td>
<td>.0378</td>
</tr>
</tbody>
</table>

(supersedes add-on `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

```
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: $\bar{y} = 6.82, s_y = 7.39, s_y^2 = 54.68 \sim 8.01\bar{y}$. 
Negative binomial distribution

- Negative binomial is a Poisson-gamma mixture.

\[ y \sim \text{Poisson} [\lambda \nu] \]
\[ \nu \sim \text{Gamma} [\mu = 1, \sigma^2 = \alpha] \]

then

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

- Probability mass function:

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

- Mean and variance:

\[
E[y] = \lambda
\]
\[
V[y] = \lambda + \alpha \lambda^2
\]

- 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

\[
\begin{align*}
E[y|x] & = \lambda = \exp(x'\beta) \\
V[y|x] & = \lambda + \alpha \lambda^2 = \exp(x'\beta) + \alpha (\exp(x'\beta))^2.
\end{align*}
\]

- The ML first-order conditions w.r.t. \( \beta \) and \( \alpha \) are (with \( \mu_i = \exp(x'_i\beta) \))

\[
\sum_{i=1}^{N} \frac{y_i - \exp(x'_i\beta)}{1 + \alpha \exp(x'_i\beta)} x_i = 0
\]

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

- Can additionally allow \( \alpha = \exp(x'\gamma) \) (generalized negative binomial).
- Can instead use Negbin 1: \( V[y|x] = (1 + \alpha)\lambda = (1 + \alpha) \exp(x'\beta) \).
- 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
Dispersion = mean                                LR chi2(7) = 773.44
Log likelihood = -10589.339                     Prob > chi2 = 0.0000

|            | Coef.   | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|------------|---------|-----------|-------|------|---------------------|
| private    | .1640928| .0332186  | 4.94  | 0.000| .0989856 - 2.292001 |
| medicaid   | .100337 | .0454209  | 2.21  | 0.027| .0113137 - 1.893603 |
| age        | .2941294| .0601588  | 4.89  | 0.000| .1762203 - 0.4120384|
| age2       | -.0019282| .0004004| -4.82 | 0.000| -.0027129 - .0011434|
| educyr     | .0286947| .0042241  | 6.79  | 0.000| .0204157 - 0.0369737|
| actlim     | .1895376| .0347601  | 5.45  | 0.000| .121409 - 0.2576662 |
| totchr     | .2776441| .0121463  | 22.86 | 0.000| .2538378 - 0.3014505|
| _cons      | -10.29749| 2.247436  | -4.58 | 0.000| -14.70238 - 5.892595|

```
/lnalpha    | -.4452773| .0306758  | -5.054007 - 3.851539
```

```
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

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

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

---

A. Colin Cameron  Univ. of Calif. - Davis  Count Regression: Part 1  Aug 28 - Sept 1, 2017  46 / 65
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 $E[y|x] = \exp(x'\beta)$
  - Then richer model can provide more efficient estimates.
- Other models imply $E[y|x] \neq \exp(x'\beta)$
  - 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 $x$ for $y \geq 1$
  i.e. only those who participate at least once are in sample.
- Truncated density (given untruncated density $f(y|x, \theta)$) is
  \[
  f(y|x, \theta, y \geq 1) = \frac{f(y|x, \theta)}{\Pr[y \geq 0|x, \theta]} = \frac{f(y|x, \theta)}{[1 - f(0|x, \theta)]}.
  \]
- 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 `zt_nb` (negative binomial)
  - `zt_nb` `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 $f(y|x, \theta)$ and cdf is $F(y|x, \theta)$)
  \[
  \begin{cases}
    f(y|x, \theta) & y \leq c - 1 \\
    1 - F(c - 1|x, \theta) = 1 - \sum_{j=0}^{c-1} f(j|x, \theta) & y = c
  \end{cases}
  \]
- Log-likelihood (where $d_i = 1$ if uncensored and $d_i = 0$ if censored)
  \[
  L(\theta) = \sum_{i=1}^{N} \left\{ d_i \ln f(y_i|x_i, \theta) + (1 - d_i) \ln(1 - \sum_{j=0}^{c-1} f(j|x_i, \theta)) \right\}
  \]
- 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
  \[ Pr[a \leq y \leq b] = \sum_{j=a}^{b} f(j|x, \theta). \]
- Let interval ranges by \([a_0, a_1 - 1], [a_1, a_2 - 1], \ldots, [a_m, a_{m+1})\), where \(a_0 = 0, a_{m+1} = \infty\).
- Let \(d_k\) be binary indicators for whether in interval \(k\) \((k = 0, \ldots, m)\).
- Then
  \[ \ln L(\theta) = \sum_{i=1}^{N} \left[ \sum_{k=0}^{m} d_{ij} \ln \left( \sum_{k=a_k}^{a_{k+1}-1} f(j|x, \theta) \right) \right]. \]
- 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 $f_1(y|x_1, \theta_1)$ so $\Pr[y = 0] = f_1(0)$ and $\Pr[y > 0] = 1 - f_1(0)$.
  - Positives: density $f_2(y|x_2, \theta_2)$ 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 ($\geq 1$)?

- Combined density is

$$f(y|x_1, x_1, \theta_1, \theta_2) = \begin{cases} 
  f_1(y|x_1, \theta_1) & y = 0 \\
  \frac{f_1(y|x_1, \theta_1)}{1 - f_1(0|x_1, \theta_1)} \times f_2(y|x_2, \theta_2) & y \geq 1
\end{cases}$$

- MLE is inconsistent if any aspect of model misspecified.
Conditional mean is now

\[ E[y|x] = \Pr[y_1 > 0|x_1] \times E_{y_2>0}[y_2 > 0, x_2]. \]

This makes marginal effects more complicated.

Example: \( f_1(\cdot) \) is logit and \( f_2(\cdot) \) is negative binomial.

Then

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

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

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

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 $f_1(y|x_1, \theta_1)$
  - Usual model for 0: realization of 0 from density $f_2(y|x_2, \theta_2)$.

- e.g. Some zeroes are mismeasurement and some are true zeros.

- Zero-inflated model has density

$$f(y|x_1, x_1, \theta_1, \theta_2) = \begin{cases} 
  f_1(0|x_1, \theta_1) + [1 - f_1(0|x_1, \theta_1)] \times f_2(0|x_2, \theta_2) & y = 0 \\
  [1 - f_1(0|x_1, \theta_1)] \times f_2(y|x_2, \theta_2) & y \geq 1
\end{cases}$$

- MLE is inconsistent if any aspect of model misspecified.
- Not used much in econometrics - hurdle model more popular.
Zero-inflated negative binomial: Stata command \texttt{zinb} and \texttt{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
Log likelihood = -10492.88

LR chi2(7) = 588.93
Pr > 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| .058957   | 4.83  | 0.000| .1691776 .4002874   |
| age2     | -.0018781| .000392   | -4.79 | 0.000| -.0026469 -.0011093|
| educyr   | .0253991| .004143   | 6.13  | 0.000| .0172786 .035196    |
| actlim   | .1737716| .033646   | 5.16  | 0.000| .1078258 .2397173   |
| totchr   | .229991 | .012079   | 19.04 | 0.000| .2063156 .2536663   |
| _cons    | -9.680235| 2.204161  | -4.39 | 0.000| -14.00031 -5.36016   |

|          | Coef.   | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|----------|---------|-----------|-------|------|----------------------|
| 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  |

Vuong test of zinb vs. standard negative binomial: z = 6.48 Pr>|z| = 0.0000

A. Colin Cameron  Univ. of Calif. - Davis  
Count Regression: Part 1  
Aug 28 - Sept 1, 2017  55 / 65
Continuous mixture models

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

- Integrating out

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

- gives $y|\lambda \sim NB [\lambda, \lambda + \alpha \lambda^2]$ if $\nu \sim Gamma[1, \alpha]$.
- Different distributions of $\nu$ 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

$$y_i \sim \text{Poisson}[\mu_{ij} = \exp(x'_{ij}\beta_j + \varepsilon_{ij})]$$

$$\beta_j = W_j\gamma + v_j$$

$$\varepsilon_{ij} \sim \mathcal{N}[0, \sigma_\varepsilon^2]$$

$$v_j \sim \mathcal{N}[0, \text{Diag}[\sigma_{jk}^2]]$$

- 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)
    
    \[
    \begin{align*}
    AIC &= -2 \ln L + 2q \\
    BIC &= -2 \ln L + qk \ln N \\
    CAIC &= -2 \ln L + q(1 + \ln)N
    \end{align*}
    \]
  - Prefer model with small AIC or BIC.
  - AIC penalty for larger model too small. Bayesian IC (BIC) better.
Compare predicted means: $E[y \mid x, \hat{\theta}]$.

Compare observed frequencies $p_j$ to average predicted frequencies

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

where $\hat{p}_{ij} = \hat{Pr}[y_i = j].$
Compare AIC, BIC for regular NB, hurdle logit/NB and zero-inflated NB.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>NBREG</th>
<th>HURDLENB</th>
<th>ZINB</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>3677</td>
<td>3677</td>
<td>3677</td>
</tr>
<tr>
<td>ll</td>
<td>-10589.3</td>
<td>-10493.2</td>
<td>-10492.9</td>
</tr>
<tr>
<td>aic</td>
<td>21196.7</td>
<td>21020.4</td>
<td>21019.8</td>
</tr>
<tr>
<td>bic</td>
<td>21252.6</td>
<td>21126.0</td>
<td>21125.3</td>
</tr>
</tbody>
</table>

Hurdle NB and ZINB are a 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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>docvis</td>
<td>3677</td>
<td>6.822682</td>
<td>7.394937</td>
<td>0</td>
<td>144</td>
</tr>
<tr>
<td>dvnbreg</td>
<td>3677</td>
<td>6.890034</td>
<td>3.486562</td>
<td>2.078925</td>
<td>41.31503</td>
</tr>
<tr>
<td>dvhurdle</td>
<td>3677</td>
<td>6.840676</td>
<td>3.134925</td>
<td>1.35431</td>
<td>31.86874</td>
</tr>
<tr>
<td>dvzinb</td>
<td>3677</td>
<td>6.838704</td>
<td>3.135122</td>
<td>.9473827</td>
<td>32.98153</td>
</tr>
</tbody>
</table>

. correlate docvis dvnbreg dvhurdle dvzinb (obs=3677)

<table>
<thead>
<tr>
<th></th>
<th>docvis</th>
<th>dvnbreg</th>
<th>dvhurdle</th>
<th>dvzinb</th>
</tr>
</thead>
<tbody>
<tr>
<td>docvis</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dvnbreg</td>
<td>0.3870</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dvhurdle</td>
<td>0.3990</td>
<td>0.9894</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>dvzinb</td>
<td>0.3983</td>
<td>0.9882</td>
<td>0.9982</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
```
Quantile regression

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

$$Q(\beta_q) = \sum_{i: y_i \geq x_i'\beta} q|y_i - x_i'\beta_q| + \sum_{i: y_i < x_i'\beta} (1 - q)|y_i - x_i'\beta_q|, \quad 0 < q < 1.$$ 

- 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.

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

- But if need to predict probabilities use a richer model.
  - Good starting point is negative binomial.
10. References: General sources that include counts

  *Regression Analysis of Count Data (RACD)*

- A. Colin Cameron and Pravin K. Trivedi (2009) 
  *Microeconometrics using Stata (MUS)*, chapter 17, Stata Press.

  *Microeconometrics: Methods and Applications (MMA)*, Cambridge Univ. Press.

- Pravin K. Trivedi and Murat Munkin (2010) 
More Specific References

- Count data models in addition to Cameron and Trivedi books:

- Generalized linear models books: