Assignment 4: Nonlinear (b11)
A. Colin Cameron U.C.-Davis
Data are at http://cameron.econ.ucdavis.edu/bgpe2011

1. Use data in file `mus10data.dta` **and 2002 data** (`keep if year02==1`)
We will do analysis similar to that in the slides, but with the two regressors `chronic` and `age`.
Note: Variable `age` is measured in tens of years!

(a) Perform Poisson regression of `docvis` on `chronic` and `age`, with heteroskedastic-robust standard errors computed. Are the regressors individually and jointly statistically significant at 5%?

(b) Use command `test` to test the hypothesis that one hundred years of aging has the same impact on doctor visits as having one (or more) chronic conditions.

(c) Use Stata add-on command `margins dydx(*)` to compute the average effect on the number of doctor visits of aging ten years.

(d) Use command `margins dydx(*), atmean` to compute the effect of aging ten years on the number of doctor visits for an individual with average values of the regressors.

(e) Use command `margins dydx(*), at()` to compute the effect of aging ten years on the number of doctor visits for a 50-year old with a chronic condition.

(f) Compare your answer in parts (c) and (d) to the average marginal effect obtained from OLS estimation of `docvis` on `chronic` and `age`.

(g) For the Poisson model, $\mathrm{E}[y|\mathbf{x}] = \exp(\mathbf{x}'\boldsymbol{\beta}) \Rightarrow \partial \mathrm{E}[y|\mathbf{x}]/\partial\mathbf{x} = \exp(\mathbf{x}'\boldsymbol{\beta})\boldsymbol{\beta} = \mathrm{E}[y|\mathbf{x}]\,\boldsymbol{\beta} \Rightarrow [\partial \mathrm{E}[y|\mathbf{x}]/\partial\mathbf{x}]/\mathrm{E}[y|\mathbf{x}] = \boldsymbol{\beta}$. Using this result, what is the estimated proportional impact of aging ten years on doctor visits?

2. If y takes only non-negative integer values and has geometric density with parameter λ then the density $f(y|\lambda)$ is

$$f(y) = \lambda^y (1+\lambda)^{-(y+1)}, \quad y = 0, 1, ..., \quad \lambda > 0.$$

Furthermore $\mathrm{E}[y] = \lambda$ and $\mathsf{V}[y] = \lambda(1+\lambda)$.
Here we introduce regressors and suppose that in the true model the parameter λ depends on regressors according to

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

where $\boldsymbol{\beta}$ is an unknown $K \times 1$ parameter vector and $\mathbf{x}_i$ is a nonstochastic $K \times 1$ vector.
The data are assumed to be independent over i.

(a) Show that the log-likelihood function for this example is

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

(b) Show that the first-order conditions for the geometric MLE can be simplify to

$$\frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \textstyle\sum_{i=1}^N \left(\frac{y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta})}{1+\exp(\mathbf{x}_i'\boldsymbol{\beta})}\right)\mathbf{x}_i = \mathbf{0}.$$

(c) Given this result, do you think that correct specification of the conditional mean will be sufficient to ensure that the MLE is consistent? Give a brief explanation.

(d) Show that

$$\mathrm{E}\left[\frac{\partial^2 Q(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}|\mathbf{x}_1, ..., \mathbf{x}_N\right] = \textstyle\sum_{i=1}^N \frac{-\exp(\mathbf{x}_i'\boldsymbol{\beta})}{1+\exp(\mathbf{x}_i'\boldsymbol{\beta})}\mathbf{x}_i\mathbf{x}_i'.$$

(e) Give the asymptotic distribution for the geometric MLE.

3. When $\gamma = h(\beta)$ where γ and β are scalars, the delta method yields $\mathrm{V}[\widehat{\gamma}] = \left(|\partial\gamma/\partial\beta_{\widehat{\beta}}\right)^2 \mathrm{V}[\widehat{\beta}]$.

(a) Hence give the formula for a 95% confidence interval for e^{β} given knowledge of $\widehat{\beta}$ and $\mathrm{se}_{\widehat{\beta}}$.

(b) Compute this for $e^{\beta_{\texttt{age}}}$ following the Poisson regression in question 1.

(c) Compare your answer to that from post-estimation command `nlcom exp(_b[age])`.

4. Stata command `ml` can be used to compute the MLE for a user-provided log-likelihood function. The following code does this for the Poisson MLE.

```
* 1. Poisson ML program lfpois defines the log-likelihood function (‘lnf’)
program define lfpois
  version 10.0
  args lnf theta1                  // theta1=x’b, lnf=lnf(y)
  tempvar lnyfact mu
  local y "$ML_y1"                 // Define y and mu so program is more readable
  generate double ‘mu’ = exp(‘theta1’)
  quietly replace ‘lnf’ = -‘mu’ + ‘y’*‘theta1’ - lnfactorial(‘y’)
end
* 2. Command ml model defines y and x, and here asks for robust se’s
ml model lf lfpois (docvis = chronic age), vce(robust)
* 3. Command ml maximize computes the estimator
ml maximize
```

Important: Note that ‘’ is left quote ‘ and right quote ’.
You need not understand the syntax in `lfpois` - just type out as is.
[Aside: `args` defines the program arguments; `tempvar` defines a temporary variable just used in this program; `local` defines a local macro; and the `args`, `tempvar` and `local` are from thereon referred to in single quotes as they are local to this program. The name of the dependent variable (in this example `docvis`) is stored in the global macro `ML_y1`. The global macro is then referred to with prefix `$`. In the program we define the local macro `y` to be text for the name of the global macro (hence the double quotes).]

(a) Run this program.

(b) Compare your results to those obtained using command `Poisson`.

5.(a) Use Stata command `ml` to estimate the MLE for the geometric regression of `docvis` on `chronic` and `age`. This is exactly the same as the program `lfpois` in question 4, except that you will need to replace the line
`quietly replace ‘lnf’ = -‘mu’ + ‘y’*‘theta1’ - lnfactorial(‘y’)`
with the corresponding expression for the log-density of the geometric (from question 2).

(b) Compare your results to those obtained using command `Poisson`.

(c) Which model do you prefer on the basis of fitted likelihood value: Poisson or geometric?

(d) Command `ml` uses the Newton-Raphson iterative method. Use your answers in question 2 to give the algebraic expressions for the Newton-Raphson method for the geometric. For simplicity use the expected Hessian $\mathrm{E}[\mathbf{H}_s]$ rather than the Hessian $\mathbf{H}_s$.