Week 10

Event Counts I

Rich Frank
University of New Orleans
October 25, 2012

- To date, we have looked at models that share a common motivation—there is a latent continuous variable of theoretical interest, y*, that underlies an observable variable *y*.
- While the observable variables are discrete, they are manifestations of a continuous variable.
- Now we turn to a different group of models motivated by a number of types of political phenomena…

Examples of count data

- Terrorist attacks in Iraq during 2006
- Jobs created in the month of September
- Homicides in New Orleans last year
- Militarized interstate disputes globally this year
- Number of articles an academic published in a given year
- Chocolate bars I ate last year
- Beers consumed by students after an MLE class

- What do we do if the discrete variable that we have is also conceptually discrete?
- We can also motivate models of event counts as latent variable models, although in a slightly different way.
- We can think of the latent process of interest as producing event counts at some underlying *rate*.

Counts in political science

- Annual number of Supreme Court appointments (King 1987)
- Number of coups d'etats in African states (Johnson et al. 1984)
- Number of terrorist incidents (Burgoon 2006)
- Alliances and number of states at war (King 1989)

- Event count data have two important characteristics:

- They are non-negative—since we cannot have negative events, all values are zero or greater.

- They are discrete—you cannot have 20.5 terrorist attacks in a month.

- Like any model, we are modeling the expected mean (μ) as a function of some independent variables, **X**, with their estimated effect **β** on that mean.

- Since counts are non-negative, the exponential function is a natural candidate for a link function.

- This means that $e^{\beta X}$ will never produce a negative value, which is a good thing because we can never have a negative expected number of events.

$$\lambda = E(Y) = e^{\beta_i x_i}$$

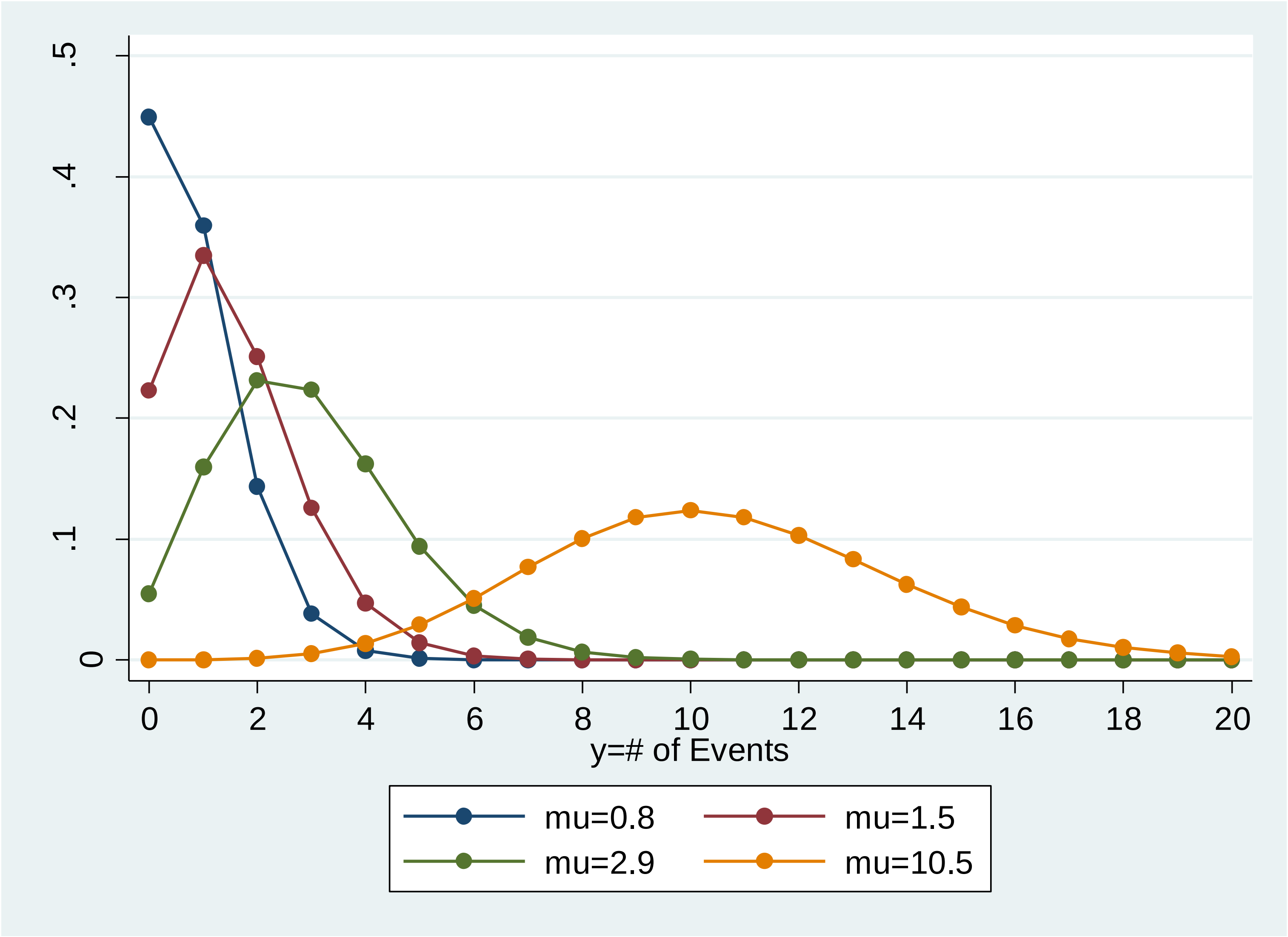
- This is about as simple of a formula as we are going to get.

Poisson distribution function

$$P(y_i = y) = \frac{e^{(-\lambda)}\mu^{y}}{y!}$$

for $y = 0, 1, 2, \ldots$

- As you can see the Poisson distribution only has one parameter λ where $\lambda > 0$.



.5
.4
.3
.2
.1
0
0
2
4
6
8
10
12
14
16
18
20
y=# of Events
mu=0.8
mu=1.5
mu=2.9
mu=10.5

```
clear
set obs 25
gen ya = .8
poisson ya, nolog
prcounts pya, plot max(20)
gen yb = 1.5
poisson yb, nolog
prcounts pyb, plot max(20)
gen yc = 2.9
poisson yc, nolog
prcounts pyc, plot max(20)
gen yd = 10.5
poisson yd, nolog
prcounts pyd, plot max(20)
label var pyapreq "mu=0.8"
label var pybpreq "mu=1.5"
label var pycpreq "mu=2.9"
label var pydpreq "mu=10.5"
label var pyaval "y=# of Events"
graph twoway connected pyapreq pybpreq pycpreq pydpreq pyaval, ///
    ytitle("Probability") ylabel(0(.1).5) xlabel(0(2)20)
graph  export 08poisson.emf, replace
```

This graph shows four important characteristics of count models (Long 1997: 218-219).

- μ is the mean of the distribution. As the mean increases the bulk of the distribution shifts to the right.
- The variance equals the mean (also known as equidispersion):

$$\text{Var(Y)} = \text{E(Y)} = \mu$$

- As μ increases, the probability of 0's decreases.
- As μ increases, the Poisson distribution approximates the normal distribution.

Important accompanying assumptions

- The observations are independent.
  - Observing one terrorist attack does not increase the probability of observing another one (!).
- There is no over-dispersion (or under-dispersion).
  - The conditional variance is not larger than the conditional mean.
  - Conditional means conditional on the values of the *X*s.
- There are no more 0's than would be predicted by the Poisson distribution.

- Once you estimate your model, you can figure out your expected baseline number of events by multiplying your coefficients by your independent variables at their means or modes.

- To estimate the ML model you want to first figure out the joint probability of your observed data.

$$\mathrm{P}(\lambda \mid \mathrm{Y}) = \prod_{i=1}^{n} F(Y \mid \lambda)$$

Then the log-likelihood is:

$$\mathrm{lnL}(\lambda \mid \mathrm{Y}) = \sum_{i=1}^{n} \ln\left(\frac{e^{(-\lambda)}\mu^{y}}{y!}\right)$$

Put another way

$$L(\beta \,|Y,X) = \prod_{i=1}^{n} P(y_i |\, \mu_i) = \prod_{i=1}^{n} \frac{e^{(-\lambda)}\mu^{y}}{y!}$$

Which we log to ease numerical maximization

$$\ln L(\beta \,|Y,X) = -n\lambda + \sum_{i=1}^{n} y_i ln\, \lambda - \sum_{i=1}^{n} \ln(y_i!)$$

- Like the ordered probit and logit the Poisson's assumptions are often restrictive given what we know about our political phenomena.
- This leads us to different models that relax these assumptions.
- More on those later…

Let's play with some actual data

- Long (1997) and Long and Freese (2006) use an example that is close to our hearts as academics—publication.

- These data from Long (1990) look at the number of publications produced by PhD biochemists.

Summary information

```
. sysuse  couart2, clear
(Academic Biochemists / S Long)

. describe

Contains data from ./couart2.dta
  obs:           915                          Academic Biochemists / S Long
 vars:             6                          30 Jan 2001 10:49
 size:        11,895 (99.9% of memory free)   (_dta has notes)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
art             byte   %9.0g                  Articles in last 3 yrs of PhD
fem             byte   %9.0g       sexlbl     Gender: 1=female 0=male
mar             byte   %9.0g       marlbl     Married: 1=yes 0=no
kid5            byte   %9.0g                  Number of children < 6
phd             float  %9.0g                  PhD prestige
ment            byte   %9.0g                  Article by mentor in last 3 yrs
-------------------------------------------------------------------------------
Sorted by:  art

. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         art |       915    1.692896    1.926069          0         19
         fem |       915    .4601093    .4986788          0          1
         mar |       915    .6622951     .473186          0          1
        kid5 |       915     .495082      .76488          0          3
         phd |       915    3.103109    .9842491       .755       4.62
-------------+--------------------------------------------------------
        ment |       915    8.767213    9.483916          0         77
```

- Let's start by just estimating the model parameter:

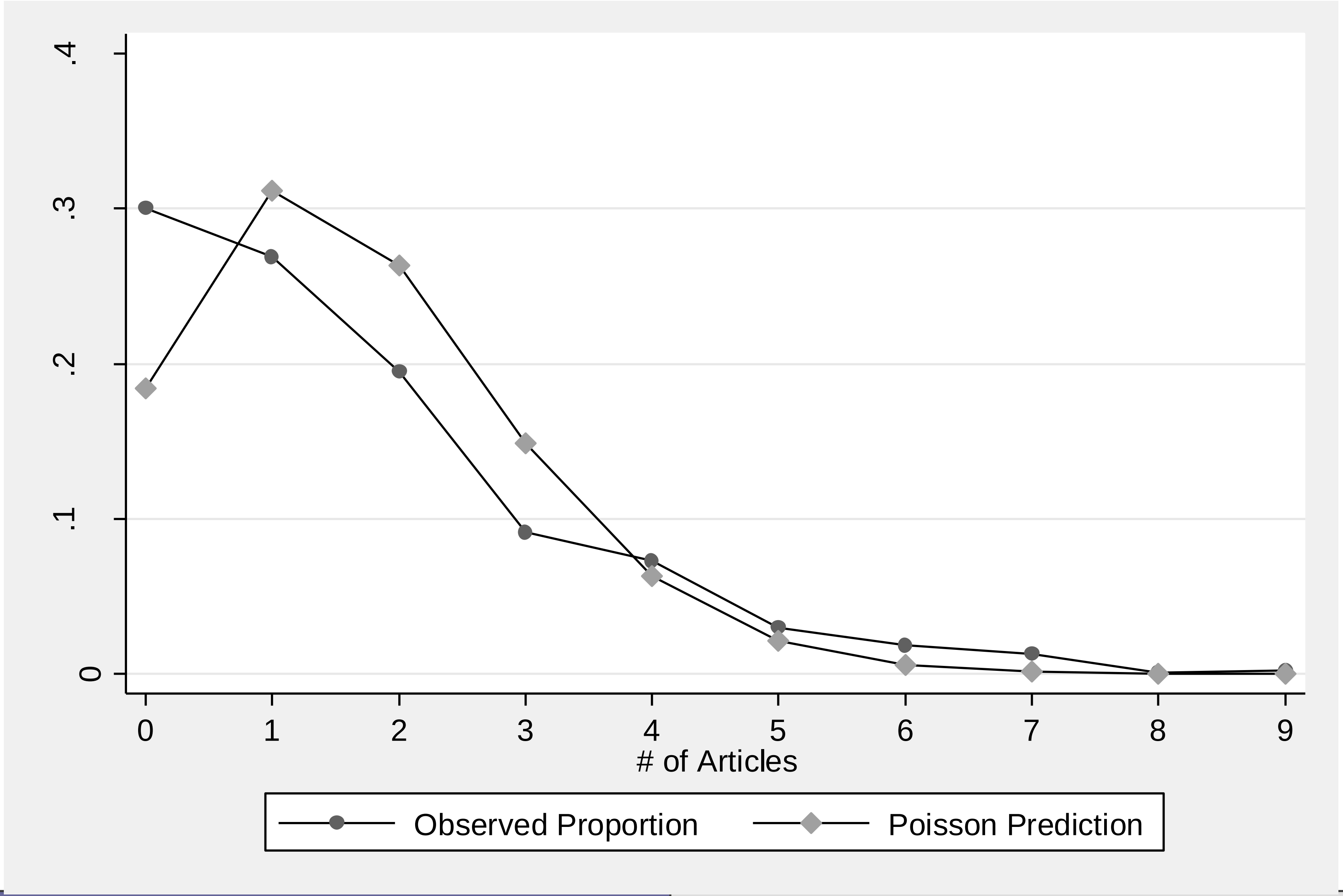
$$\mu = e^{\beta_0}$$

```
. poisson art, nolog
Poisson regression                                Number of obs    =        915
                                                  LR chi2(0)       =       0.00
                                                  Prob > chi2      =          .
Log likelihood = -1742.5735                       Pseudo R2        =     0.0000

------------------------------------------------------------------------------
         art |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .5264408   .0254082    20.72   0.000     .4766416      .57624
------------------------------------------------------------------------------
```

- Therefore the $\hat{\mu} = e^{(.5264408)} = 1.6929$

```
. prcounts psn, plot max(9)

. label var psnobeq "Observed Proportion"

. label var psnpreq "Poisson Prediction"

. label var psnval "# of Articles"

. list psnval psnobeq psnpreq in 1/10

     +-------------------------------+
     | psnval    psnobeq    psnpreq |
     |-------------------------------|
  1. |      0   .3005464   .1839859 |
  2. |      1   .2688525    .311469 |
  3. |      2   .1945355   .2636423 |
  4. |      3   .0918033    .148773 |
  5. |      4    .073224   .0629643 |
     |-------------------------------|
  6. |      5   .0295082   .0213184 |
  7. |      6   .0185792    .006015 |
  8. |      7   .0131148   .0014547 |
  9. |      8   .0010929   .0003078 |
 10. |      9   .0021858   .0000579 |
     +-------------------------------+
```

Comparing observed and predicted



Stata Commands:

```
graph twoway connected psnobeq psnpreq psnval, ///
    ytitle("Probability") ylabel(0(.1).4) ///
    xlabel(0(1)9) ysize(2.7051) xsize(4.0421)
graph export 08psnpred.emf, replace
```

- As you can see the Poisson model under-predicts 0's and over-predicts 1-3.
- More on this later…

A basic Poisson model

```
. sysuse  couart2, clear
(Academic Biochemists / S Long)

. poisson art fem mar kid5 phd ment, nolog

Poisson regression                                Number of obs   =        915
                                                  LR chi2(5)      =     183.03
                                                  Prob > chi2     =     0.0000
Log likelihood = -1651.0563                       Pseudo R2       =     0.0525

------------------------------------------------------------------------------
         art |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         fem |  -.2245942   .0546138    -4.11   0.000    -.3316352   -.1175532
         mar |   .1552434   .0613747     2.53   0.011     .0349512    .2755356
        kid5 |  -.1848827   .0401272    -4.61   0.000    -.2635305   -.1062349
         phd |   .0128226   .0263972     0.49   0.627     -.038915    .0645601
        ment |   .0255427   .0020061    12.73   0.000     .0216109    .0294746
       _cons |   .3046168   .1029822     2.96   0.003     .1027755    .5064581
------------------------------------------------------------------------------
```

- You interpret the results depending on if you are interested in the expected value of *y* or the distribution of *y*.

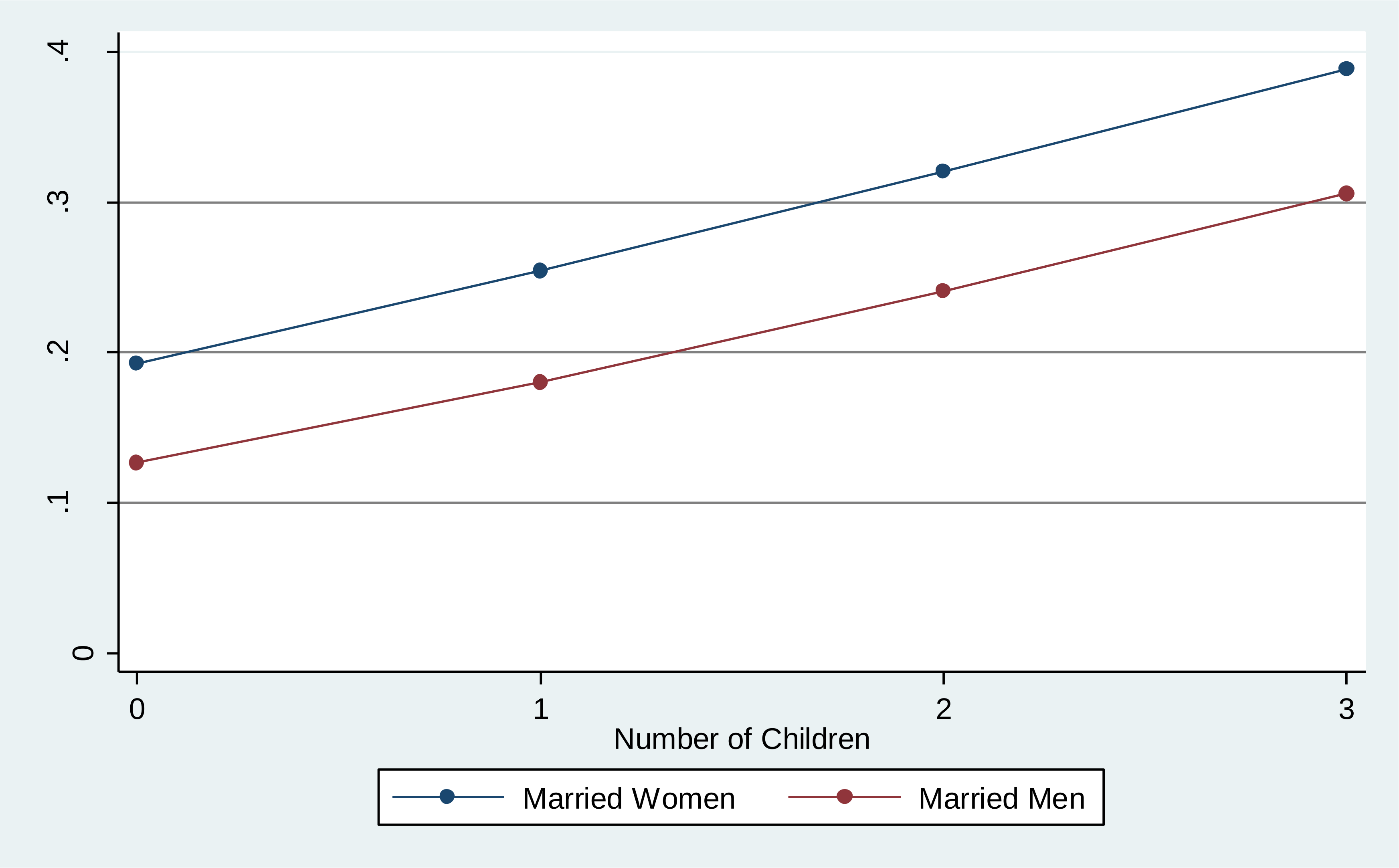
- Interpretation is much more straightforward than multinomial logit.

$$\mu = E(y|x) = e^{\beta X}$$

I’m often partial to percent change

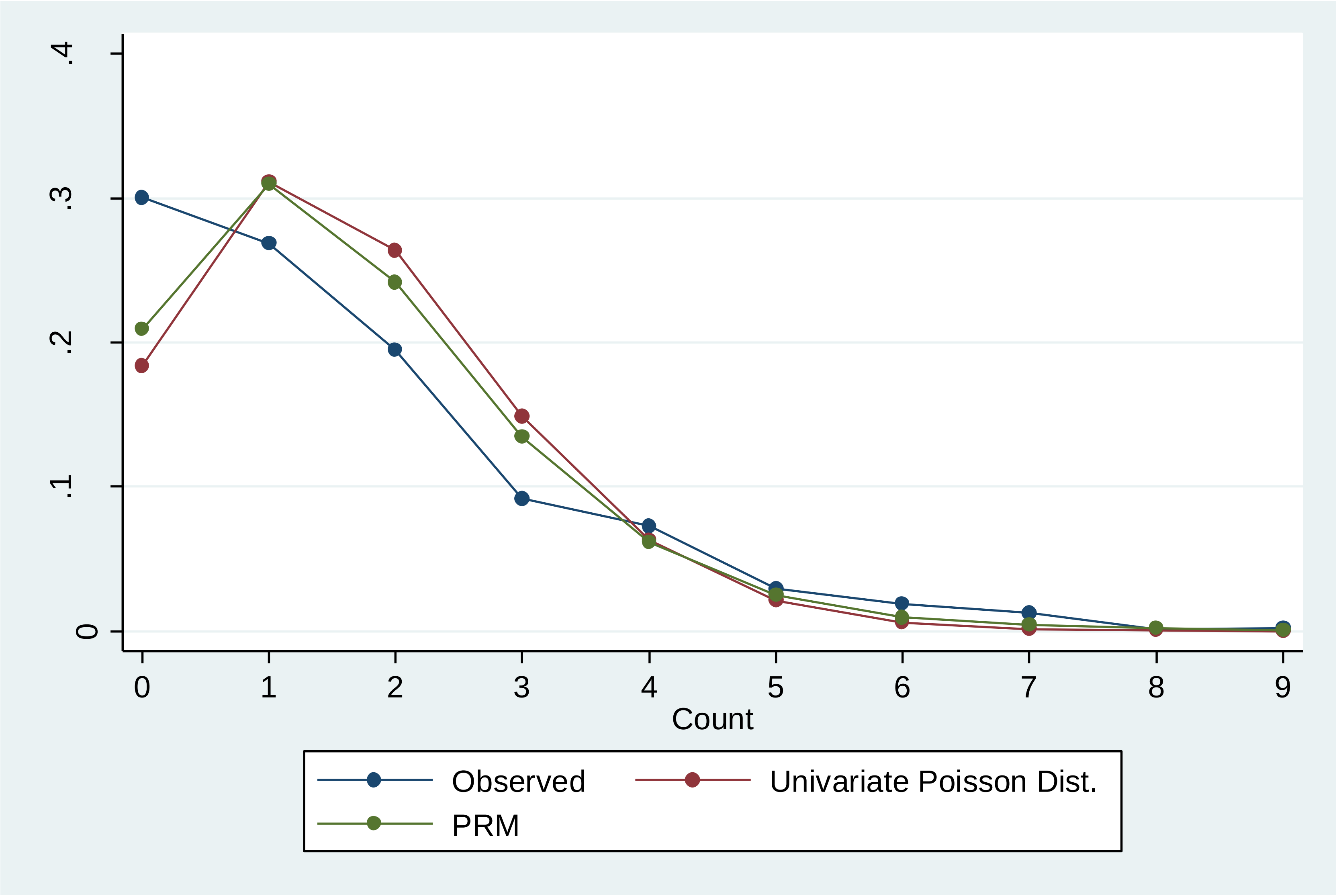
```
. * percent change
. listcoef fem ment, percent help

poisson (N=915): Percentage Change in Expected Count

 Observed SD: 1.926069

----------------------------------------------------------------------
         art |        b          z      P>|z|        %       %StdX      SDofX
-------------+--------------------------------------------------------
         fem |  -0.22459    -4.112    0.000      -20.1     -10.6     0.4987
        ment |   0.02554    12.733    0.000        2.6      27.4     9.4839
----------------------------------------------------------------------
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
       % = percent change in expected count for unit increase in X
   %StdX = percent change in expected count for SD increase in X
   SDofX = standard deviation of X
```

Using predicted probabilities



```
. * prgen
. prgen kid5, x(fem=1 mar=1) rest(mean) from(0) to(3) gen(fprm) n(4)

poisson: Predicted values as kid5 varies from 0 to 3.

           fem        mar       kid5        phd       ment
x=           1          1  .49508197  3.1031093  8.7672131

. prgen kid5, x(fem=0 mar=1) rest(mean) from(0) to(3) gen(mprm) n(4)

poisson: Predicted values as kid5 varies from 0 to 3.

           fem        mar       kid5        phd       ment
x=           0          1  .49508197  3.1031093  8.7672131

. label var fprmp0 "Married Women"

. label var mprmp0 "Married Men"

. label var mprmx  "Number of Children"

.
. * graph predictions
. graph twoway connected fprmp0 mprmp0 mprmx, ///
>     ylabel(0(.1).4) yline(.1 .2 .3) xlabel(0(1)3) ///
>     ytitle("Probability of No Articles") ///
>     ysize(2.5051) xsize(4.0421)
```

Difference between significant and important predictors



```
. prcounts prm, plot max(9)

. label var prmpreq "PRM"

. label var prmobeq "Observed"

. graph twoway connected prmobeq psnpreq prmpreq prmval, ///
>     ytitle("Probability of Count") ylabel(0(.1).4) ///
>     xlabel(0(1)9) ysize(2.7051) xsize(4.0413)
```

- As you can see the output and the interpretation of Poisson models are straightforward and similar to what we have done previously.

- However, in practice most scholars are not able to use Poisson because their data violate the rather restrictive assumptions of the Poisson model.

- The most frequently violated assumption is that of equidispersion.

- What do we do if we have overdispersion?
- We turn to the negative binomial regression model (NBRM or NB).
- The NB model estimates an additional parameter, α, that reflects *unobserved* heterogeneity among observations.

- The NB model adds an error, ε, that is assumed to be uncorrelated with the X's.

- We then estimate a random variable $\tilde{\mu}$:

$$\tilde{\mu}_i = e^{(\beta x_i + \varepsilon_i)} = \mu_i e^{\varepsilon_i} = \mu_i \delta_i$$

Where $\delta_i = e^{\varepsilon_i}$

- In order to be able to identify our model (like we do with the logit/probit, and multinomial) we assume a value for the mean of the error term.

- The most convenient assumption is that $\mathrm{E}(\delta_i) = 1$.

- This lets us have the same expected count as the Poisson despite allowing a new source of variation.

$$\mathrm{E}(\tilde{\mu}_i) = \mathrm{E}(\mu_i \delta_i) = \mu_i E(\delta_i) = \mu_i$$

- To compute P(Y | X, δ) rather than P(Y | X) we need to specify a distribution for δ in addition to the distribution for Y.
- The most common distribution for δ is the gamma function (which is part of the normal distribution used in the probit).

$$g(\delta_i) = \frac{{v_i}^{v_i}}{\Gamma(v_i)} {\delta_i}^{v_i - 1} e^{(-\delta_i v_i)} \text{ for } v_i > 0$$

As v_i gets bigger the distribution becomes more bell shaped centered on 1.

- The negative binomial distribution is given by the formula in Long (1997: 232).

- The expected value of Y in the NB distribution is the same as the Poisson:

$$E(y|x) = \mu_i$$

- But the variance is different and given by:

$$\text{Var}(y|x) = \mu_i\left(1 + \frac{\mu_i}{v_i}\right) = e^{\beta x_i}\left(1 + \frac{e^{\beta x_i}}{v_i}\right)$$

- As you might be able to guess, the Poisson is nested within the NB when $\alpha = 0$.

- Again, to estimate we assume that *v* is the same for all individuals.

- Long (1997: 234-5) plots sample data and shows the difference in errors between the Poisson and the NB.

What causes the unobserved heterogeneity?

- Contagion?
- Unobserved heterogeneity via omitted variables?

- The ML log-likelihood equation is in Long (1997: 236).
- Let’s get back to the article publishing example.

```
. nbreg art fem mar kid5 phd ment, nolog

Negative binomial regression                      Number of obs   =        915
                                                  LR chi2(5)      =      97.96
Dispersion     = mean                             Prob > chi2     =     0.0000
Log likelihood = -1560.9583                       Pseudo R2       =     0.0304

------------------------------------------------------------------------------
         art |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         fem |  -.2164184   .0726724    -2.98   0.003    -.3588537   -.0739832
         mar |   .1504895   .0821063     1.83   0.067    -.0104359    .3114148
        kid5 |  -.1764152   .0530598    -3.32   0.001    -.2804105     -.07242
         phd |   .0152712   .0360396     0.42   0.672    -.0553652    .0859075
        ment |   .0290823   .0034701     8.38   0.000     .0222811    .0358836
       _cons |    .256144   .1385604     1.85   0.065    -.0154294    .5277174
-------------+----------------------------------------------------------------
    /lnalpha |  -.8173044   .1199372                     -1.052377   -.5822318
-------------+----------------------------------------------------------------
       alpha |   .4416205   .0529667                      .3491069    .5586502
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) =  180.20 Prob>=chibar2 = 0.000
```

- See the estimate of the ln(α) and α?
- ln(α) is used to force α estimates to be positive. The estimated α is given on the next line.
- How do the estimates differ between Poisson and NB?

Variable	PRM	NBRM
Gender: 1=female 0=male	-0.225	-0.216
	-4.11	-2.98
Married: 1=yes 0=no	0.155	0.150
	2.53	1.83
Number of children < 6	-0.185	-0.176
	-4.61	-3.32
PhD prestige	0.013	0.015
	0.49	0.42
Article by mentor in last 3 yrs	0.026	0.029
	12.73	8.38
Constant	0.305	0.256
	2.96	1.85
alpha		0.442
N	915	915

legend: b/t

- Notice that the *z*-values for the NB are substantially smaller than those for the Poisson.

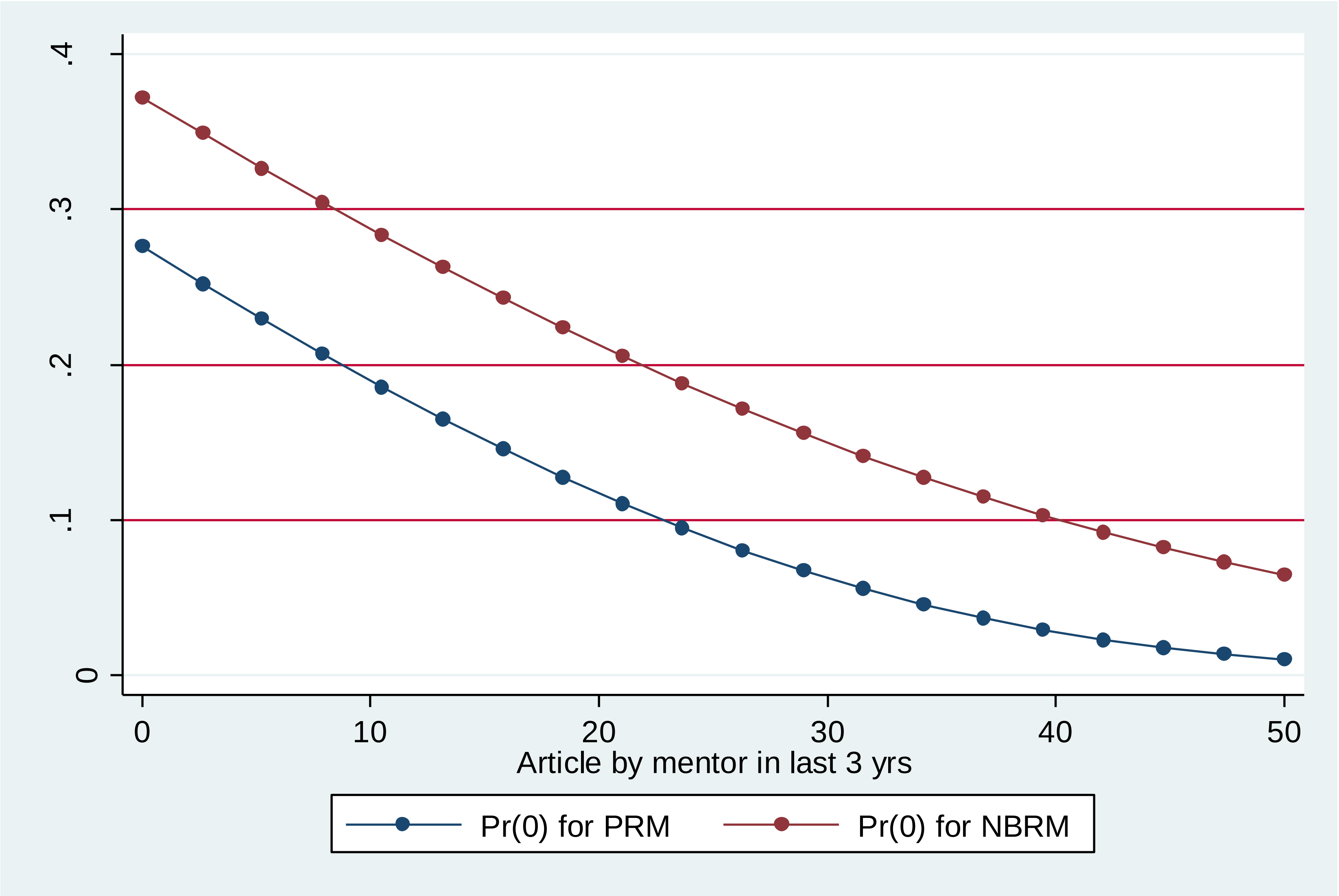
- To test $H_0{:}\ \alpha = 0$ Stata does a likelihood ratio test:

$$G^2 = 2(lnL_{NB} - lnL_{Poisson})$$

Where G^2 is distributed chi-squared.

- The results in the last slide therefore suggest that there is significant evidence of overdispersion. This means that the NB is preferred over the Poisson.

- Interpretation of the NB is the same as for the Poisson, except that the predicted probability formula is not just $e^{\beta X}$ but by a much more complicated formula (formula 8.17 in Long 1997: 237).



.4
.3
.2
.1
0
0
10
20
30
40
50
Article by mentor in last 3 yrs
Pr(0) for PRM
Pr(0) for NBRM

- This helps explain why we might better prefer the NB to the Poisson.
- Remember that the Poisson under-predicted the 0’s in the initial analysis.
- The NB better matches predictions to the actual data.

- However, the sheer number of zeros raises some other concerns that we will deal with in more detail next week.

```
. tab art

Articles in |
 last 3 yrs |
     of PhD |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        275       30.05       30.05
          1 |        246       26.89       56.94
          2 |        178       19.45       76.39
          3 |         84        9.18       85.57
          4 |         67        7.32       92.90
          5 |         27        2.95       95.85
          6 |         17        1.86       97.70
          7 |         12        1.31       99.02
          8 |          1        0.11       99.13
          9 |          2        0.22       99.34
         10 |          1        0.11       99.45
         11 |          1        0.11       99.56
         12 |          2        0.22       99.78
         16 |          1        0.11       99.89
         19 |          1        0.11      100.00
------------+-----------------------------------
      Total |        915      100.00
```

- Now, let’s turn to the King articles and Gowa (1998) to look at these models in the wild as well as how they were introduced to the field.