POLS 7050

Spring 2008

April 9, 2007

Models for Event Count Data, II

Heterogeneity, Contagion, and Overdispersion–Oh My!
Event count models are similar to some of our earlier models (e.g. logits and probits) in that they can be characterized as a realization of a latent process. In the event count case, what is latent is the *rate* at which events occur.

An example I like to use is watching for animals. Suppose that, having nothing better to do, we decide to count the number of cats that wander through our backyard in a day (or Emus or Walruses depending on where you live). Over two weeks, we observe $Y_{cats} = \{0, 1, 1, 0, 2, 0, 1, 0, 3, 1, 2, 1, 0, 2\}$; this yields $\widehat{Y} = 1.0, \sigma_{cats} = 0.92$.

Now suppose there is some underlying rate at which the events occur, such that we'd expect some number of cats to pass through on a particular day; call this rate $\lambda_c ats$. We're interested in the probability that we observe $\{0, 1, 2, 3, ...\}$ cats per day. To figure out this probability, we might make four assumptions about the process generating the events (cats):

1. Zero events have occurred at the beginning of the period.
2. More than one event can't occur at the same time (in the cat case, this may be a bit hard to swallow, but in other examples, it isn't...)
3. The periods are all of the same length. (This actually isn't all that critical, as we discussed earlier...).
4. *The probability of an event occurring is constant within a particular period, and independent of other events during the same period.* This assumption is critical, as we'll soon see...

If these assumptions hold, then the number of events observed in a particular period is a Poisson process:

$$Pr(Y_t = y) = \frac{\exp(-\lambda)\lambda^y}{y!} \tag{1}$$

for $\lambda > 0$ and $y \in \{0, 1, 2, ...\}$. The parameter λ is the unobserved "rate" of occurrence; it is also the expected value of the variable Y [i.e., $E(Y) = \lambda$]. We usually use an exponential "link" function to allow the mean to vary according to some set of independent variables [$i.e., \lambda_i = \exp(X_i \boldsymbol{\beta})$]. This yields the log-likelihood:

$$\ln L = \sum_{i=1}^{N} \{-\exp(\mathbf{X}_i \boldsymbol{\beta}) + Y_i \mathbf{X}_i \boldsymbol{\beta} - \ln(Y_i!)\} \quad (2)$$

where the last term is often (but not always) dropped, since it doesn't vary in $\boldsymbol{\beta}$

The last, two-part assumption we made is critical to what the event count data generated will "look like"...

- The first assumption is about the *independence* of events.
- The second assumption requires a *constant rate* of event occurrences.

We'll consider each of these in turn...

Independence and Contagion: Antelope and Foxes

Antelope

Now suppose that, instead of cats, we're observing antelope. Antelope are herd animals: where you see one, you'll likely see more. This fact suggests that counts of antelope will probably violate the independence assumption, that the occurrence of one event has no effect on the likelihood of observing additional events in the same period. For antelope, one event increases the likelihood of another; this is an example of **positive contagion**.

What's the effect of this positive contagion? The answer is that we'll have greater numbers of higher and low counts...

- We might go nearly two weeks without seeing any, then see seven each on the last two days (when a herd wanders through our neighborhood)
- So, we'd observe $Y_{antelope} = \{0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 7\}$.
- This count is also $\bar{Y}_{antelope} = 1.0$, but has $\sigma_{antelope} = 6.46$.

The point: *Positive contagion increases the variance of the observed counts*, even if it does not affect the mean; this is known as **overdispersion**.

But, we already know that for the Poisson, $\text{E}(Y) = \text{Var}(Y) = \lambda$. So, if we fit a Poisson model to the antelope data, we'll be imposing the (incorrect) mean-variance equality restriction on the estimation. As a result,

- We'll effectively be requiring the variance to be less than it really is.
- The consequence is that we will underestimate the true variability in the data.
- This will lead us to underestimate our standard errors, and so to overestimate the degree of precision in our coefficients.

Foxes

Now suppose we're observing foxes instead. Foxes are territorial animals; seeing one fox means its unlikely you'll see another (different) fox any time soon. This means that the occurrence of one event *decreases* the probability of another event in the same period (and so also violates the independence assumption). This phenomenon is called **negative contagion**...

What's the effect of negative contagion?

- Negative contagion will yield greater numbers of counts right around the mean; this is known as **underdispersion**.
- E.g., in counting foxes for two weeks, we might see: $Y_{foxes} = \{1, 0, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1\}$.
- Y_{foxes} also has $\bar{Y}_{foxes} = 1.0$, but has $\sigma_{foxes} = 0.15$ – that is, *less* variability than for an independent (Poisson) process.

Fitting a Poisson model to these data imposes $\text{E}(Y) = \text{Var}(Y) = \lambda_{foxes}$. This means that doing so will artificially *overestimate* the variability in Y_{foxes}, and lead us to overestimate the standard errors, and underestimate our precision/confidence in the model parameters.

The larger point is that the mean-variance equality restriction of the standard Poisson distribution is a direct result of the assumptions we make ab out the process generating the events.

Cross-Period Effects

Over and above the issue of event arrivals within observation periods, note as well that, if our "time periods" are arbitrary, then contagion (positive or negative) *within* observations also implies contagion *across* observations. As a result – and by the same logic – contagion across observations can also lead to over- or under-dispersed data, relative to the Poisson.

As an example, consider the data above if we aggregate our counts to two-day periods. We then have data that look like:

$$
\begin{aligned}
Y_{cats} &= \{1, 1, 2, 1, 4, 3, 2\} \\
Y_{antelope} &= \{0, 0, 0, 0, 0, 0, 14\} \\
Y_{foxes} &= \{1, 2, 2, 3, 2, 2, 2\}
\end{aligned}
$$

Again, the mean of each variable is the same (2.0), but the variances are very different. Moreover, we can think of the contagion that (previously) was "within" a period, and now see that it also operates "across" periods.

Heterogeneity

The discussion of event dependence ignores the second Poisson event assumption: that of constant event arrival rates. The assumption that rates are constant implies that rates are *uniform within time periods*; i.e., that all micro-"events" have equal probability. If the observed count is made up of aggregates of multiple units, this is unlikely to be the case.

For example, Wawro talks about heterogeneity in counts of Presidential vetoes during a Congressional session. If a President is more likely to veto bills (say) early in a session (when s/he can do so to make political hay) than later in the session (when appropriations, etc. need to be passed), then the assumption of a constant within-period rate λ is violated.

The assumption of constant rates also gives rise to a second (implicit) belief about the data: that, conditional on any covariates $\mathbf{X}$, the value of λ is the same *across observations*. Put differently, this means that (in a regression context) all of the systematic influences on the event arrival rate λ are included in $\mathbf{X}\boldsymbol{\beta}$, so that only (constant) random noise remains. We term this *unobserved heterogeneity*, and can think of this as equivalent to:

- Assuming that the model is "correctly" specified (i.e., as a specification matter), or
- Assuming that the non-systematic variance in Y is truly "random" (that is, an assumption about the implicit "errors"), or
- Assuming that the conditional (on $\mathbf{X}$ and $\boldsymbol{\beta}$) expectation of Y is constant.

Unobserved heterogeneity of this type also leads to overdispersion (that is, $\text{E}(Y) < \text{Var}(Y)$), of a form exactly the same as that for positive contagion outlined above. Intuitively, this is because the non-constant rate induces greater random variability in Y than would a constant λ. To illustrate this, suppose we have a model like:

$$\lambda_i \equiv \text{E}(Y_i) = f[\mathbf{X}_i\boldsymbol{\beta} + Z_i\theta] \tag{3}$$

and where we fail to include Z as a relevant covariate. So long as $\theta \neq 0$, the result will be that some observations will have different (conditional on $\mathbf{X}$) values of λ than will others – that is, that the data will contain heterogeneity.

Summing Up

Formally, recall that the Poisson model requires that $\text{E}(Y_i) = \text{Var}(Y_i) = \lambda_i$. Consider relaxing the mean-variance equality restriction, by saying that $\text{Var}(Y) = \lambda_i\sigma$. Then, in general, we can think of three situations:

$$\begin{aligned}
\text{Poisson Dispersion} &\leftrightarrow \text{E}(Y) = \text{Var}(Y) \leftrightarrow \sigma = 1\\
\text{Overdispersion} &\leftrightarrow \text{E}(Y) < \text{Var}(Y) \leftrightarrow \sigma > 1\\
\text{Underdispersion} &\leftrightarrow \text{E}(Y) > \text{Var}(Y) \leftrightarrow 0 < \sigma < 1
\end{aligned}$$

Generally, political science data are not Poisson distributed, even conditionally. We typically study dependent processes (vs., say, engineers or physicists), and we regularly fail to measure stuff, leading to unobserved heterogeneity. Or, at best, we measure it badly (and, in fact, measurement error in right-hand-side variables also yields overdispersion – see Prentice 1986).

Dealing with Over- and Underdispersion

Overdispersion

Overdispersion is much more common in political science settings (and, in fact, in nearly all others as well) than is underdispersion. Moreover, in most instances where we have covariates $\mathbf{X}_i$, what we really care about is *conditional* over- or underdispersion; that is, over- or underdispersion in the *errors/residuals*.

A Test for Overdispersion

It seems we ought to be able to test for overdispersion pretty easily – just estimate a Poisson model, and then see whether the (squared) "errors" have a variance that is statistically different from λ_i.

Formally, this test is just a t-test for $\hat{\delta} = 0$ in the equation:

$$\hat{u}_i = \delta\hat{\lambda}_i + \epsilon_i \tag{4}$$

where

$$\hat{u}_i = \frac{(Y_i - \hat{\lambda}_i)^2 - Y_i}{\hat{\lambda}_i\sqrt{2}} \tag{5}$$

and $\hat{\lambda}_i$ is the predicted value of λ_i for observation i (that is, $\exp(\mathbf{X}_i\hat{\boldsymbol{\beta}})$ from the Poisson regression). This test thus has three steps:

1. Estimate a Poisson regression of Y_i on $\mathbf{X}_i$, and generate predicted counts $\hat{\lambda}_i$.
2. Calculate $\hat{u}_i$ according to (5), above.
3. Estimate (4) using OLS, and test $H_0 : \hat{\delta} = 0$.

We'll illustrate an example of this in a bit. For now, I should mention that there are a bunch of other tests for overdispersion, including LM and Wald tests based on the negative binomial model (see below). Cameron and Trivedi (1998, §3.4 and 5.6) gives a good discussion of these.

Models for Overdispersed Counts

Intuitively, to address the idea of heterogeneity, we can just drop the assumption that the rate λ_i is constant within an observation, and instead make it a random variable. The result is that we have the usual conditional Poisson variate, but with a random error term. Formally, we simply specify that

$$\begin{aligned} \mathrm{E}(Y_i) \equiv \lambda_i &= \exp(\mathbf{X}_i\boldsymbol{\beta} + u_i) \\ &= \exp(\mathbf{X}_i\boldsymbol{\beta})\exp(u_i) \\ &= \lambda_i \nu_i \end{aligned} \tag{6}$$

With this approach, we then have to specify the distribution of u_i in order to identify the model. We could (in theory) use a lot of different distributions, but we usually use the Gamma, for two reasons:

1. Is a natural one for variability, since its nonnegative, and

2. It leads to a nice (if complicated) closed-form solution.

If the νs in (6) are assumed to be randomly distributed according to a one-parameter Gamma distribution with mean $\mathrm{E}(\nu) = 1$ and variance $\mathrm{Var}(\nu) \equiv \sigma^2 = \frac{1}{\alpha}$, then the marginal density of Y is *negative binomial*:

$$\Pr(Y_i = y|\lambda_i, \alpha) = \left(\frac{\Gamma(\alpha^{-1} + Y_i)}{\Gamma(\alpha^{-1})\Gamma(Y_i + 1)}\right)\left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i}\right)^{\alpha^{-1}}\left(\frac{\lambda_i}{\lambda_i + \alpha^{-1}}\right)^{Y_i} \tag{7}$$

where Γ is the gamma function:

$$\Gamma(a) = \int_0^\infty \exp(-t)t^{a-1}dt$$

As before, we again typically model $\lambda_i = \exp(\mathbf{X}_i\boldsymbol{\beta})$. This model has $\mathrm{E}(Y) = \lambda$, just like the Poisson, but $\mathrm{Var}(Y) = \lambda(1 + \alpha\lambda)$, $\alpha > 0$.

- Thus, the variance is allowed to be greater than the mean, *but...*

- ... the variance is still (positively) dependent on the mean (i.e., heteroscedastic), as we would hope for an event-count variate.

- Intuitively, larger values of α correspond to greater amounts of overdispersion.

Note as well that the model reduces to the Poisson when $\alpha = 0$. Cameron and Trivedi call this the "NB2" model, since it can also be expressed as $\mathrm{Var}(Y) = \lambda + \alpha\lambda^2$. With $\lambda_i = \exp(\mathbf{X}_i\boldsymbol{\beta})$, we get the log-likelihood:

$$\ln L_{NB} = \sum_{i=1}^{N}\left\{\left(\sum_{j=0}^{Y_i-1}\ln(j+\alpha^{-1})\right) - \ln Y_i! - (Y_i-\alpha^{-1})\ln[1+\alpha\exp(\mathbf{X}_i\boldsymbol{\beta})] + Y_i\ln\alpha + Y_i\mathbf{X}_i\boldsymbol{\beta}\right\} \quad (8)$$

To summarize: If we assume an event count process with gamma heterogeneity in the rate of event arrivals, the resulting event count will follow a *negative binomial* distribution. This means that:

- The variance is always larger than the mean.
- When $\alpha = 0$, then $\mathrm{E}(Y) = \mathrm{Var}(Y)$ and the negative binomial model reduces to the Poisson.
- Because α is restricted to be greater than 0, most programs (e.g. Stata) actually estimate either $\frac{1}{\alpha}$ or $\ln(\frac{1}{\alpha})$.
- This also suggests an easy likelihood-ratio test for overdispersion: Simply estimate Poisson and negative binomial models, and then calculate $-2 \times (\widehat{\ln L_{Poisson}} - \widehat{\ln L_{NB}})$; this test is distributed χ^2_1; larger values of the test reject the null hypothesis of no overdispersion. (Note as well that a t-test of $\hat{\alpha} = 0$ will, asymptotically, give the same results).

Finally, the model remains log–linear; moreover, we still have $\widehat{E(Y_i)} \equiv \hat{\lambda}_i = \exp(\mathbf{X}_i\hat{\boldsymbol{\beta}})$. This means that the IRR and predicted count approaches to interpretation remain open to us. Likewise, while we probably wouldn't want to do it "by hand," we can also calculate predicted probabilities for discrete counts, conditional on $\mathbf{X}_i$ and $\hat{\boldsymbol{\beta}}$ (Stata and most other software will do this for us...).

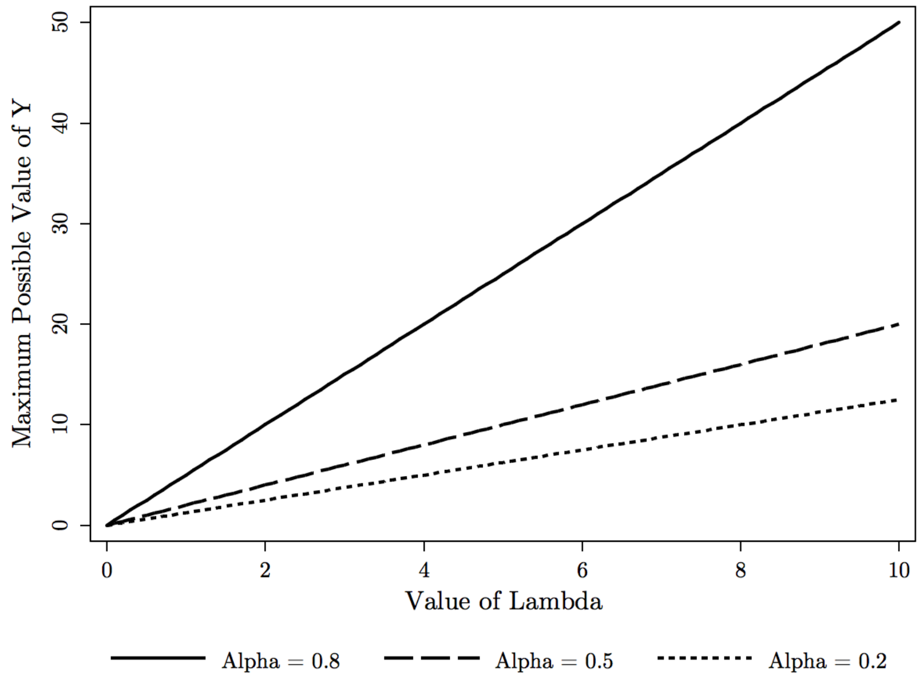
Underdispersion

A similar type of distribution to the NB, the "continuous parameter binomial" (CPB), is used to model underdispersed data (that is, data that have *negative contagion*). Not surprisingly, this is a variant of the binomial distribution which is "scaled" to insure that its probabilities sum to 1.0. The density for the CPB model is:

$$\Pr(Y_i = y|\lambda_i, \alpha) = \frac{\frac{\Gamma\left(\frac{-\lambda_i}{\alpha-1}+1\right)}{Y_i!\Gamma\left(\frac{-\lambda_i}{\alpha-1}-Y_i+1\right)}(1-\alpha)^{Y_i}(\alpha)^{\frac{-\lambda_i}{\alpha-1}-Y_i}}{D_i} \quad (9)$$

where D_i is the aforementioned scaling factor and is equal to nothing more than the sum from 0 to $\frac{-\lambda_i}{\alpha-1}+1$ of the binomial distribution, above. This model:

Figure 1: Maximum Possible Values of Y as a Function of λ, for Three Values of α in the CPB Distribution.



- also has $\mathrm{E}(Y_i) = \lambda_i$ [again, typically with $\mu_i = \exp(\mathbf{X}_i\boldsymbol{\beta})$], but
- has $\mathrm{Var}(Y) = \lambda_i\alpha$ with $0 < \alpha < 1$. In addition,
- it also reduces to the standard Poisson when $\alpha = 1$.

Note also that the CPB distribution imposes a theoretical "upper limit" on the count variable. In particular,

$$\max(Y_i) = \frac{-\lambda_i}{\alpha - 1}. \tag{10}$$

This limit is due to the fact that the variability of Y is constrained by α; as $\alpha \to 1.0$, the upper limit disappears (i.e., goes to infinity). Figure 1 presents potential maximum values of Y as a function of λ for varying values of α in the CPB distribution.

The log-likelihood for the CPB model above is:

$$\begin{aligned}\ln L_{CPB} &= \sum_{i=1}^{N}\left\{\ln\Gamma\left(\frac{-\lambda_i}{\alpha-1}+1\right) - \ln\Gamma\left(\frac{-\lambda_i}{\alpha-1} - Y_i + 1\right)\right. \\ &\left. +Y_i\ln(1-\alpha) + \left(\frac{-\lambda_i}{\alpha-1} - Y_i\right)\ln(\alpha) - \ln(D_i)\right\} \end{aligned} \tag{11}$$

Practical Stuff

In practice, underdispersed data are pretty rare (though not as rare as some would have you believe). Also, for Stata users, I should mention that Stata will estimate negative binomial models, but not CPB ones.

Another point that bears repeating is that, as I noted above, when speaking of over– or underdispersion, we usually mean conditional on the effects of the independent variables [i.e. $\text{Var}(Y_i|\mathbf{X}_i, \boldsymbol{\beta})$]. When we introduce covariates, we are factoring out some of the heterogeneity in the data (that is, it is no longer unobserved, but now part of the systematic part of the model). So, as model specification gets better, we often see our data go from overdispersed, to Poisson, to underdispersed. (There's a nice example of this in the Stata 9 manuals, under the heading `-poisson-`).

A Quick Illustrative Example

Consider some data on Supreme Court decisions during the Warren and Burger Courts (1953-1985 Terms). We'll act like we have data on three variables:

- `namici` is the number of amicus curiae briefs filed in each case,
- `term` is the term (i.e., year) of the court,
- `civlibs` is whether (=1) or not (=0) the case involved a civil rights and liberties issue.

The data look like this:

```
. summm namici term civlibs

    Variable |     Obs        Mean   Std. Dev.       Min        Max
-------------+-----------------------------------------------------
      namici |    7161    1.025136   2.544066          0         53
        term |    7157    71.12114   9.194212         53         85
     civlibs |    7161    .5009077   .5000341          0          1
```

The mean count is a bit over 1.0, but there are also *lots* of zeros (in fact, nearly 69 percent of the cases had no briefs filed at all) and a very small number of cases with high numbers of briefs; this is consistent with overdispersion (note that the variance of the raw counts is 6.47, substantially higher than its mean). Our expectation is that both `term` and `civlibs` will have a positive influence on the number of amicus briefs filed in each case.

We'll start with a Poisson model:

```
. poisson namici term civlibs

Poisson regression                                Number of obs   =       7157
                                                  LR chi2(2)      =    2199.42
                                                  Prob > chi2     =     0.0000
Log likelihood = -13427.959                       Pseudo R2       =     0.0757
------------------------------------------------------------------------------
  namici |     Coef.   Std. Err.     z    P>|z|    [95% Conf. Interval]
---------+--------------------------------------------------------------------
    term |  .0636112      .00147  43.27  0.000      .06073    .0664923
 civlibs | -.2979656    .0234971 -12.68  0.000    -.344019   -.2519122
   _cons | -4.511961    .1119035 -40.32  0.000   -4.731288   -4.292634
------------------------------------------------------------------------------

. poisson, irr

Poisson regression                                          Number of obs   =       7157
                                                            LR chi2(2)      =    2199.42
                                                            Prob > chi2     =     0.0000
Log likelihood = -13427.959                                 Pseudo R2       =     0.0757

------------------------------------------------------------------------------
      namici |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        term |   1.065678   .0015666    43.27   0.000     1.062612    1.068753
     civlibs |   .7423269   .0174425   -12.68   0.000     .7089154     .777313
------------------------------------------------------------------------------
```

These results indicate that, at least in these data,

- the average numbers of *amicus* briefs filed in Supreme Court cases *increased* during the 1953-1985 period, but
- cases involving civil rights and liberties issues actually have *fewer amicus* briefs filed than do other sorts of cases.

Looking at the incidence rate ratios, the latter effect says that each term saw an average increase in the expected number of *amicus* briefs of about six-and-a-half percent, and that civil liberties cases have only about 74 percent of the expected number of *amici* filed as do other cases.

We can implement the test for overdispersion outlines in (4) “by hand”:

```
. predict Poissonhat
(option n assumed; predicted number of events)
(4 missing values generated)

. gen uhat = ((namici - Poissonhat)^2 - namici) / (Poissonhat * (sqrt(2)))
(4 missing values generated)

. regress uhat Poissonhat

      Source |       SS       df       MS              Number of obs =    7157
-------------+------------------------------           F(  1,  7155) =    6.14
       Model |  4959.86151     1  4959.86151           Prob > F      =  0.0132
    Residual |  5778317.78  7155  807.591584           R-squared     =  0.0009
-------------+------------------------------           Adj R-squared =  0.0007
       Total |  5783277.65  7156  808.171834           Root MSE      =  28.418

------------------------------------------------------------------------------
        uhat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  Poissonhat |   1.465829    .591486     2.48   0.013     .3063417    2.625316
       _cons |   1.579188   .6934805     2.28   0.023     .2197611    2.938615
------------------------------------------------------------------------------
```

Here, if the data are conditionally Poisson-distributed, the estimated coefficient on `Poissonhat` should equal zero. We can reject that hypothesis quite confidently (at around $p = .01$), and the fact that the estimate is greater than zero tells us that overdispersion is likely in the data. Accordingly, we would next estimate a negative binomial model:

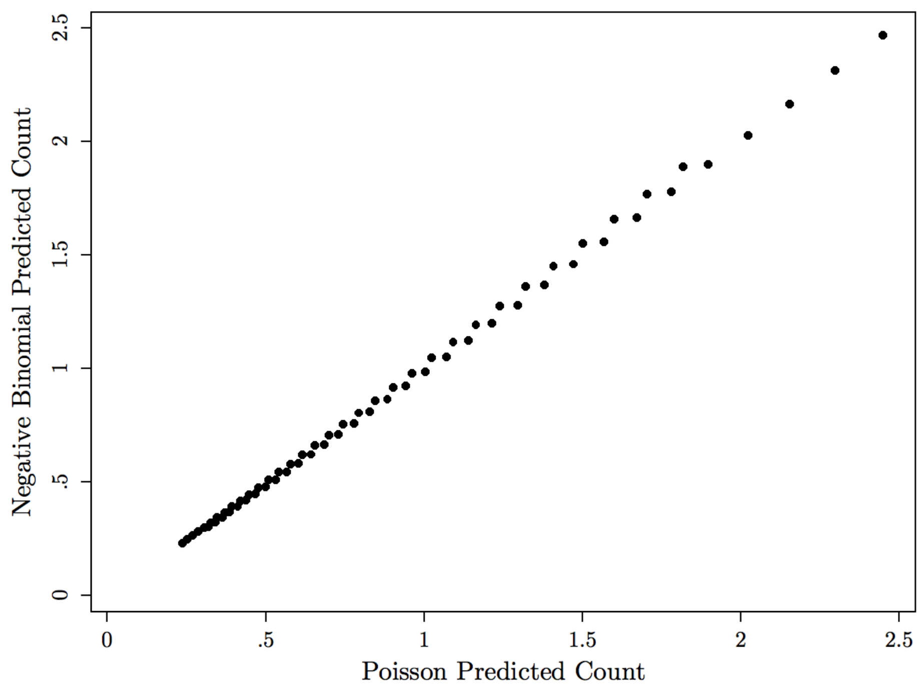
```
. nbreg namici term civlibs

Negative binomial regression                      Number of obs   =       7157
                                                  LR chi2(2)      =     441.86
                                                  Prob > chi2     =     0.0000
Log likelihood = -8685.0839                       Pseudo R2       =     0.0248
--------------------------------------------------------------------
  namici |     Coef.   Std. Err.    z    P>|z|  [95% Conf. Interval]
---------+----------------------------------------------------------
    term |  .0657306   .0030845  21.31   0.000    .059685    .0717761
 civlibs | -.2677686   .0538811  -4.97   0.000 -.3733737   -.1621635
   _cons | -4.683137   .2242108 -20.89   0.000 -5.122582   -4.243692
---------+----------------------------------------------------------
```

```
/lnalpha |  1.360013   .0326324                      1.296054    1.423971
---------+----------------------------------------------------------------
   alpha |  3.896243   .1271439                      3.654848    4.153583
--------------------------------------------------------------------------
LR test of alpha=0:   chibar2(01) = 9485.75    Prob>=chibar2 = 0.000
```

Note a few things about these results:

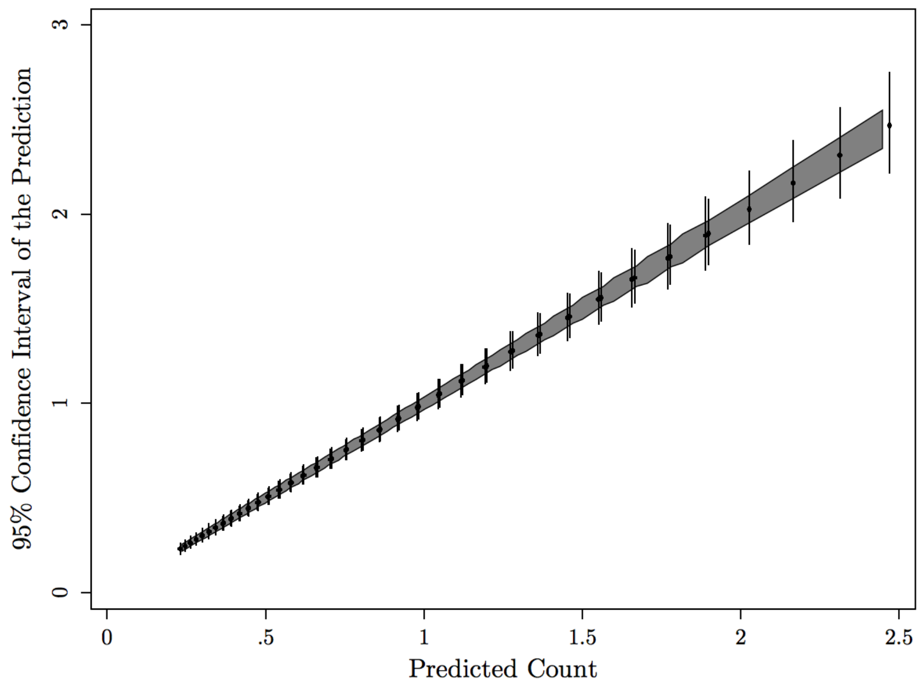
- Stata estimates $\ln(\alpha)$, rather than α; this is because $\alpha = 0$ is a boundary condition. (Because one can think of α as the variance of the individual-specific heterogeneity parameter ν_i, α is necessarily greater than or equal to zero; it can never be negative).
- The NB model strongly indicates overdispersion – the estimate $\hat{\alpha}$ is 3.89, vastly different from zero. The LR test – which, given the boundary condition, is a better test to use than a t-test – clearly rejects the Poisson in favor of the NB.
- Despite all that, the coefficient estimates for $\hat{\beta}$ don't change all that much, nor do the model predictions (see Figure 2).

Figure 2: Predicted $\hat{Y}$s: Poisson vs. Negative Binomial



- What does change are the standard errors – those for the NB model are 2-3 times those for the Poisson. This is consistent with the fact that, in the presence of overdispersion, the Poisson model will tend to underestimate one's standard errors. This last fact can be seen in the estimated standard errors and confidence intervals for the predictions (Figure 3), which are substantially larger for the negative binomial predictions (the solid bars) than for the Poisson model (the shaded area).

Figure 3: Predicted 95% Confidence Intervals: Poisson vs. Negative Binomial



Taking it a step further, we might want to see if the extent of overdispersion changes over time, or if there is more or less overdispersion in civil liberties cases than in others. To do this, we'll estimate a variance-function negative binomial model:

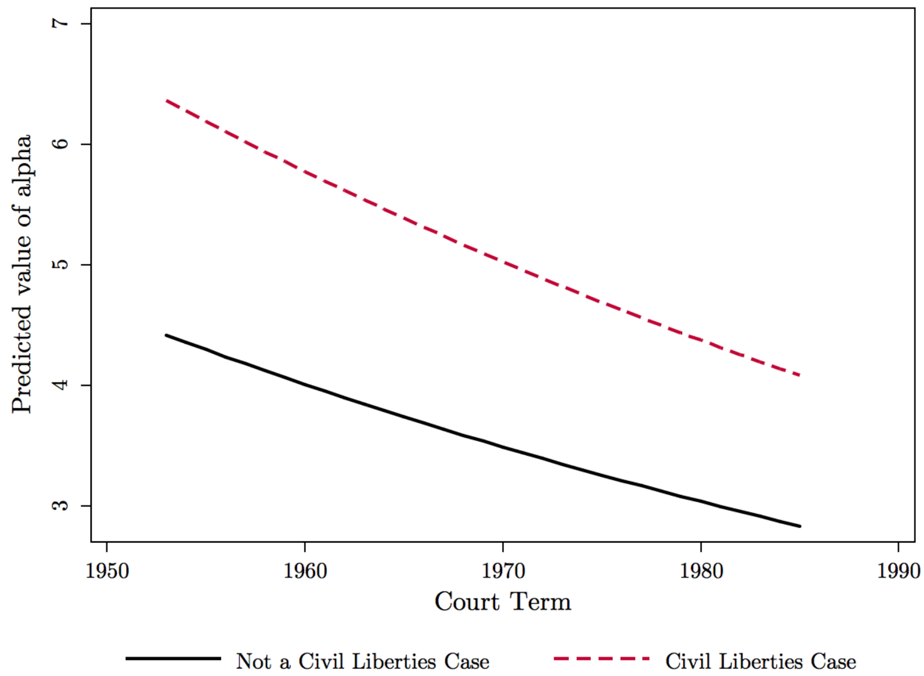
```
. gnbreg namici term civlibs, lna(term civlibs)

Generalized negative binomial                 Number of obs   =       7157
   regression                                 LR chi2(2)      =     401.80
                                              Prob > chi2     =     0.0000
Log likelihood = -8664.5025                   Pseudo R2       =     0.0227
------------------------------------------------------------------------------
  namici |      Coef.  Std. Err.     z    P>|z|    [95% Conf. Interval]
---------+--------------------------------------------------------------------
namici   |
    term |   .0658922  .0031151   21.15   0.000    .0597868    .0719977
 civlibs |  -.2747653  .0545366   -5.04   0.000    -.381655   -.1678755
   _cons |  -4.691456  .2290076  -20.49   0.000   -5.140302   -4.242609
---------+--------------------------------------------------------------------
lnalpha  |
    term |  -.0138852  .0040337   -3.44   0.001   -.0217912   -.0059792
 civlibs |   .3652729  .0649148    5.63   0.000    .2380422    .4925035
   _cons |   2.221755  .3030342    7.33   0.000    1.627819    2.815691
------------------------------------------------------------------------------
```

Note a few things about this model:

- An LR test versus the single-parameter NB model yields $-2(-8685.08-(-8664.50)) = 41.16$, which is $\sim \chi^2_2$ and yields $p < .0001$. This suggests that the "generalized" model fits significantly better than the single–parameter NB model.
- We also find that
  - The `term` variable has a negative effect on the dispersion of the event count – this might be because (e.g.) we do a "better job" of explaining event counts later in the data (i.e., there is less unobserved heterogeneity in later terms) or because there is less positive contagion in later terms – perhaps because allied groups are coordinating more.
  - Conversely, `civlibs` has a positive effect on the variance of the counts – civil liberties cases have higher variances (that is, greater extra–Poisson variance) than do other sorts of cases.

Both of these findings are illustrated in Figure 4, which plots the (out-of-sample) predicted conditional variance $\hat{\alpha}$ for both civil liberties and non-civil liberties cases across the range of values for term. (In theory, one could use `-predict ..., stdplna-` to get standard errors – and therefore confidence intervals – around these predictions, but for some reason that command is not working. I have an e-mail in to the Stata folks to find out what's up with that...).

Figure 4: Predicted Values of $\hat{\alpha}$, by `term` and `civlibs`



Finally, a key point is the following: This example is a badly underspecified model, so we're on dangerous ground attributing the variance effects to any one mechanism. In general, the better specified one's model, the less heterogeneity – and, therefore, overdispersion – one will have to worry about.