Multichotomous Dependent Variables III

1 Introduction

Last time we looked at unordered or nominal multichotomous dependent variables. We are now going to look at event count models.

1. Ordered Dependent Variables:
   - Presidential Approval - Approve, Indifferent, Disapprove
   - Political Interest - very, somewhat, not much, not at all
   - International Conflict - war, diplomatic conflict, peace.
2. Unordered Dependent Variables
   - Which of three parties did you vote for?
   - Did you travel by car, plane or train?
   - Which potential government coalition formed?
3. Count Dependent Variables
   - Number of Wars
   - Number of Government Formation Attempts
   - Number of Court Cases Heard

It is important to recognize that a couple of things that might look like event count data are not event count data!

- **Ordinal Data**

  This is what we looked at last week - you should use something like ordered probit or ordered logit.

- **Grouped Binary Data**

  Data which are the number of successes or failures out of some known number of binary trials is known as grouped binary data. For example, the number of successful veto overrides in each Congress is an example of grouped binary data. Although grouped binary data can be expressed as event counts, they are not event counts. As a result, you should generally not use event count models for these data. The exception is when there are relatively few successes relative to the possible number of trials. Grouped binary data can be evaluated with a binomial or an extended beta-binomial model. For more information on this, see King (2001, 117-121).

Event counts have certain characteristics - they are discrete and non-negative. For this reason, using OLS for event count data can be bad. King (1988) explains how OLS is wrong and why. The bottom line is that OLS produces inaccurate estimates (could produce negative counts) and is inefficient (because they fail to take account of the heteroskedastic nature of event counts).

2 Poisson Model

The most basic model for event counts is the Poisson model. There are two ways of coming up with the Poisson model as a good model of event counts.[1]

2.1 Approach 1

The first approach is to start with an abstract model of event counts that assumes (i) events occur over time; (ii) there is a constant rate (λ) at which events occur - this rate is the expected number of events in period of length h; (iii) events are independent in that the occurrence of one event does not alter the probability of another; and (iv) as the length of the interval h goes to zero the probability of an event occurring in the interval $(t, t+h] = \lambda h$ and the probability of two events occurring in the interval is zero. Our dependent variable is the number of events that occur in interval t of length h. The probability that the number of events occurring in $(t, t+h]$ is equal to some value $y \in \{0, 1, 2, \dots\}$ is:

$$\Pr(Y_t = y_i) = f(y_i) = \frac{e^{-\lambda_i h} \lambda h^{y_i}}{y_i!} \tag{1}$$

In this Poisson process, events occur independently with a constant probability equal to λ times the length of the interval i.e λh. Typically, we assume that all intervals are of the same length equal to one, and so we have

$$\Pr(Y_t = y_i) = f(y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \tag{2}$$

This is the form of the Poisson distribution that you are most likely to see.

2.2 Approach 2

The second way of coming up with the Poisson model is to think of events counts as counts of rare events. Specifically, a Poisson random variable approximates a binomial random variable when the binomial parameter n (number of trials) is large and p (probability of a success) is small. Remember that the binomial distribution represents the probability of y successes (events) in N independent trials (possible events), where the probability of an event in each trial is p, has the following distribution:

$$P(Y = y) = \binom{N}{y} p^y (1-p)^{N-y} \tag{3}$$

[1] Much of this material is directly based on notes from Chris Zorn.

where $\binom{N}{y} = \frac{N!}{(N-y)!\, y!}$. I've dropped subscripts at this point.

We can think of the binomial distribution in terms of time i.e. think of the number of trials as a length of time. For example, we can divide a span of time of length 1 into N 'sub-intervals' of length $\frac{1}{N}$. The probability of an event (p) equals $\frac{\lambda}{N}$ where λ is simply the number of events that occurred in all of the N periods. Thus, we can rewrite the binomial as

$$P(Y = y) = \binom{N}{y} \left(\frac{\lambda}{N}\right)^y \left(1 - \frac{\lambda}{N}\right)^{N-y} \tag{4}$$

By dividing the span of time into more and more intervals and making the probability of an event (p) proportionally smaller, the time span becomes a continuum. Now we want to know what the probability of y events is as $N \rightarrow \infty$. This is:

$$\begin{aligned} P(Y = y) &= \lim_{N\rightarrow\infty} \binom{N}{y} \left(\frac{\lambda}{N}\right)^y \left(1 - \frac{\lambda}{N}\right)^{N-y} \\ &= \frac{e^{-\lambda}\lambda^y}{y!} \end{aligned} \tag{5}$$

This is the same as Eq. (2). The bottom line is that for a large number of Bernoulli trials, where the probability of an event in any one trial is small, the total number of events observed will follow a Poisson distribution.

2.3 Characteristics

The Poisson distribution has certain characteristics.

1. As λ increases, the mass of the distribution shifts to the right. This is because $E[y] = \lambda$. λ can be thought of as the rate or expected number of times an event occurs per unit of time (fits with the binomial derivation) or as the expected or mean count (fits with our more abstract model of event counts).
2. Var(y) $= E[y] = \lambda$. This means that the model is restrictive (as is any model with only one parameter).
3. As λ increases the probability of 0s decreases.
4. As λ increases, the Poisson distribution approximates a normal distribution.

2.4 Estimation

As always, we need to parameterize λ_i as $f(x_i, \beta)$. We know that λ must be positive because it is a count i.e. $E(Y) > 0$. As a result, we usually assume that:

$$E(Y_i) = \lambda_i = e^{x_i \beta} \tag{6}$$

This yields the following probability model:

$$\begin{aligned} Pr(Y_i = y) &= \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \\ &= \frac{e^{-e^{x_i\beta}} (e^{x_i\beta})^{y_i}}{y_i!} \end{aligned} \tag{7}$$

Thus, the likelihood function for the whole sample is just:

$$\begin{aligned} \mathcal{L} = &= \prod_{i=1}^{N} \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \\ &= \prod_{i=1}^{N} \frac{e^{-e^{x_i\beta}} (e^{x_i\beta})^{y_i}}{y_i!} \end{aligned} \tag{8}$$

The log likelihood is:

$$\begin{aligned} \ln\mathcal{L} &= \sum_{i=1}^{N} [-\lambda_i + y_i x_i \beta - \ln(y_i!)] \\ &= \sum_{i=1}^{N} [-e^{x_i\beta} + y_i x_i \beta - \ln(y_i!)] \end{aligned} \tag{9}$$

Note that we can essentially ignore the last term $\ln(y_i!)$ when maximizing the log-likelihood since it does not vary with λ_i and, hence, with β. The log-likelihood is globally concave and hence easy to estimate. The gradient vector is

$$\begin{aligned} G = \frac{\partial \ln\mathcal{L}}{\partial \beta} &= \sum_{i=1}^{N} -e^{x_i\beta} x_i + y_i x_i \\ &= \sum_{i=1}^{N} \left(y_i - e^{x_i\beta} \right) x_i \\ &= \sum_{i=1}^{N} \left(y_i - \lambda_i \right) x_i = 0 \end{aligned} \tag{10}$$

The Hessian matrix is

$$H = \frac{\partial^2 \ln\mathcal{L}}{\partial\beta\partial\beta} = -\left(x_i e^{x_i\beta}\right)' x_i$$
$$= -\left(x_i\lambda_i\right)' x_i \tag{11}$$

To estimate a Poisson model in STATA, you simply type:

```
poisson Y X
```

2.5 Interpretation

The following example uses data on presidential veto overrides from Charles Franklin. In the U.S. government, presidents can veto legislation passed by congress. This prevents the bill from becoming law unless two-thirds of both the House and Senate vote to 'override' the veto. If two-thirds of both houses agree, the bill becomes law despite the President's objections; otherwise the bill dies. We are going to treat the number of vote overrides as an event count which may follow the Poisson or some other distribution.[2] We are going to assume that it follows the Poisson distribution for now. I base the Poisson model on the following set of independent variables.

$$\begin{aligned} x_i\beta = \beta_0 &+ \beta_1\text{MarginHouse} + \beta_2\text{MarginSenate} \\ &+ \beta_3\text{CongressExperience} + \beta_4\text{GubernatorialExperience} \\ &+ \beta_5\text{Reelect} + \beta_6\text{PresidentVote} \end{aligned} \tag{12}$$

where MARGINHOUSE and MARGINSENATE represent the margin that the president's party has in the House and Senate, CONGRESSEXPERIENCE and GUBERNATORIALEXPERIENCE indicate whether the president has congressional or gubernatorial experience, REELECT is whether the president is up for reelection, and PRESIDENTVOTE is the percentage of the popular vote that the president won in the last election. Results are shown in Table 1.

2.5.1 Coefficients

It is possible to use the sign of the coefficients in Table 1 to make inferences about the effect of the independent variables on the expected number of counts of veto overrides. For example the larger the margin controlled by the president's party in the Senate, the lower the expected number of veto overrides (coefficient is negative). The more congressional and gubernatorial experience that the president has, the higher the expected number of veto overrides (coefficients are positive). We should obviously go further in analyzing our results if we want to talk about substantive significance.

[2] It may be more appropriate to treat these kind of data as grouped binary data.

Table 1: The Determinants of the Number of Presidential Veto Overrides

Dependent Variable: Number of Presidential Veto Overrides Regressor	 Poisson
MarginHouse	0.004
	(0.005)
MarginSenate	-0.06***
	(0.02)
CongressExperience	1.90**
	(0.74)
GubernatorialExperience	1.97**
	(0.79)
Reelect	0.46
	(0.35)
PresidentVote	0.05*
	(0.03)
Constant	-4.33**
	(1.72)
Log likelihood	-35.41
Observations	26

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-tailed)
Standard errors are given in parentheses.

2.5.2 Expected Count

We could calculate the expected number of veto overrides for a particular situation:

$$\lambda = E[y] = e^{X_i\beta} \tag{13}$$

I do this in Table 2 where I calculate the expected count for Scenario 1: MARGINHOUSE=Mean, MARGINSENATE=Mean, CONGRESSEXPERIENCE=0, GUBERNATORIALEXPERIENCE=0, REELECT=1, and PRESIDENTVOTE=Mean. Scenario 2 differs in that GUBERNATORIALEXPERIENCE=1. As expected from the results in Table 1, if the president has gubernatorial experience there is likely to be more presidential vetos overriden (1.98 on average) than if the president does not have gubernatorial experience.

Table 2: Expected Count and First Difference in Counts

	Strongly Disapprove
Expected Count	0.43
(Scenario 1)	[0.06, 1.40]
Expected Count	2.41
(Scenario 2)	[1.02, 4.83]
Difference	1.98
	[0.51, 4.32]

Notes: 95% Confidence interval in parentheses

2.5.3 Marginal Effects

We could calculate marginal effects of the independent variables on the expected number of counts:

$$\frac{\partial E[y]}{\partial x_k} = e^{X_i\beta}\beta_k = E[y]\beta_k = \lambda\beta_k \tag{14}$$

The marginal effect of increasing MARGINSENATE in scenario 1 above is

$$\begin{aligned}\frac{\partial E[y]}{\partial \text{MarginSenate}} &= \left(e^{-4.33+0.004\times -9.73+-0.06\times 2.5+1.90\times 0+1.97\times 0+0.46\times 1+0.05\times 52.64}\right)\times -0.06\\ &= \left(e^{-1.43}\right)\times -0.06\\ &= 0.24\times -0.06\\ &= -0.014\end{aligned} \tag{15}$$

While we could calculate marginal effects, the nonlinearity of the model suggests that this might not be a quantity that we are all that interested in - this is the same as with the ordered probit and logit. First differences might be better.

2.5.4 Incident Rate Ratios

We could also calculate what is called the Incident Rate Ratio or Factor Change. The quantity of interest in the Poisson model tends to be the expected count, $\hat{\lambda}$. So we could compare two observations by dividing the expected count for one scenario by the expected count of a different scenario. For example:

$$\frac{E[y|X, x_k+\delta]}{E[y|X, x_k]} = e^{\beta_k\delta} \tag{16}$$

where δ is some substantively interesting change in one of the independent variables. Note that the effect of a change in x_k does not depend on the level of any of the variables. If we wanted to know the effect of the president having gubernatorial experience, we would simply calculate $e^{1.97\times 1} = 7.17$. Thus, the expected count increases by a factor of 7.17 if the president has gubernatorial experience. Put another way, the incident rate of presidential veto overrides is 7.17 times higher when the president has gubernatorial experience compared to an identical situation when he does not. You can get these incident rate ratios automatically from STATA:

```
poisson Y X, irr
```

You can also calculate the percentage change in the expected count for a δ change in an independent variable by using the following equation:

$$\text{Percentage Change} = 100[e^{\beta_k\delta} - 1] \tag{17}$$

Continuing with our particular example, this means that there is a $100[7.17 - 1] = 617\%$ increase in the expected count of presidential veto overrides when the president has gubernatorial experience compared to when the president does not have gubernatorial experience.

2.5.5 Predicted Probabilities

We could also calculate predicted probabilities that a particular observation takes on a particular count value y:

$$\Pr(y_i = y|x_i\hat{\beta}) = \frac{e^{-e^{x_i\hat{\beta}}}\left[e^{x_i\hat{\beta}}\right]^y}{y!} \tag{18}$$

This just requires plugging in values of substantive interest. For example, the predicted probabilities associated with Scenario 1 above are

$$\begin{aligned}\Pr(y_i = 0|\text{Scenario 1}) &= \frac{e^{-0.24} \times 0.24^0}{0!} \\ &= \frac{0.79 \times 1}{1} \\ &= 0.79\end{aligned} \tag{19}$$

$$\begin{aligned}\Pr(y_i = 1|\text{Scenario 1}) &= \frac{e^{-0.24} \times 0.24^1}{1!} \\ &= \frac{0.79 \times 0.24}{1} \\ &= 0.19\end{aligned} \tag{20}$$

$$\begin{aligned}\Pr(y_i = 2|\text{Scenario 1}) &= \frac{e^{-0.24} \times 0.24^2}{2!} \\ &= \frac{0.79 \times 0.06}{2} \\ &= 0.002\end{aligned} \tag{21}$$

and so on. If you add up these predicted probabilities you get 0.982, which tells you that these outcomes account for most of the potential outcomes at this level of covariates.

2.6 Exposure and Offsets

Having interpreted the results from the Poisson model in Table 1, we might ask ourselves whether this is the correct specification. Remember that in the Binomial (rare events) formulation of the Poisson model, we assumed that N (the number of trials or the number of possible events) went to infinity. Similarly, the general version of the Poisson model does not place an upper limit on the number of possible events. However, in our case of presidential vetoes overridden, we have an upper limit. Congress can only override a presidential veto if the president actually vetoes. Thus, the maximum number of vetoes that can be overridden if the president makes 9 vetoes is 9. This number refers to an observation's 'exposure'. King (2001, 124-126) addresses the exposure issue from a slightly different angle. He says that we can think of the exposure issue (the number of possible events when N is not infinite) as being an unequal observation interval issue i.e. each observation of the number of event counts occurs over a different interval of time - recall that so far we have assumed that each of our observations of event counts come from intervals of time that have the same length. The bottom line is that if you have a maximum number of event counts that can happen in a particular interval of time OR if your observations of event counts come from intervals of time that vary in length, then you have an 'exposure' issue to deal with.

The problem in these situations is that each observation does not have the same exposure i.e. $M_i \neq M_j, \ \forall i \neq j$. As Zorn points out, though, the expected count (λ) will be proportional to the exposure i.e. $E[y_i|x_i, M_i] = \lambda_i M_i$. Since M_i (the total number of possible vetoes that can be overridden in our case) might generally influence the expected count, λ_i, we want to divide out this value. We could do this by having:

$$\frac{E[y_i]}{M_i} = \frac{\lambda_i}{M_i} = e^{x_i\beta} \tag{22}$$

instead of $E[y_i] = e^{x_i\beta}$ as we had before. This can be rewritten as:

$$\lambda_i = e^{x_i\beta} M_i$$

Since $M_i = e^{\ln M_i}$, we can rewrite this as:

$$\lambda_i = e^{x_i\beta + \ln M_i} \tag{23}$$

The implication of this formulation is that we should include the natural log of the number of presidential vetoes (this is the natural log of the maximum number of possible overrides) on the right hand side and constrain its coefficient to be one. There are two ways of doing this in STATA.

- poisson nover hmargin smargin congexpr govexpr reelect popvote if nveto>0, exposure(nveto)
- poisson nover hmargin smargin congexpr govexpr reelect popvote, offset(lnveto)

where LNVETO=LN(NVETO).

Results from these models are shown in Table 3 below. The first column shows the results from our basic Poisson model. The second column shows the results when we take account of exposure using either STATA's 'exposure' or 'offset' commands. There are several differences between columns 1 and 2. The effect of the size of the president's party's margin in the senate becomes less significant, and the magnitude of the effect of the president having gubernatorial experience increases significantly.[3]

The results in column 2 come from a constrained model in which the coefficient on the natural log of presidential vetoes is constrained to be one. An alternative way to proceed is simply to include the natural log of presidential vetoes on the right hand side of the model and then test to see if the appropriate coefficient is one. Results from such a model are presented in column 3. The χ^2 statistic from a Wald test that the

[3] The number of observations falls by one because one observation had 0 presidential vetoes and it is not possible to take the natural log of 0. One way around this would be to add 0.001 (i.e. something small) to all observations and take the natural log of that.

coefficient on $\ln(\text{PRESIDENTVETOS}) = 1$ is 1.12 and the p-value is 0.29. Thus, I cannot reject the null hypothesis that the coefficient $= 1$ and I might simply go with the results from the exposure or offset model in column 2. However, as Zorn notes, the coefficient on ln(PRESIDENTVETOS) may be of substantive interest. For example, we might want to know how the expected number of veto overrides is affected by the number of override opportunities that they have. For example, a coefficient significantly less than one might indicate that as the number of override opportunities increases, the expected number of overrides actually decreases.

As a practical matter, Zorn notes that if (i) all the observations have the same exposures or (ii) no observation comes close to experiencing its maximum possible number of events, then the issue of exposure is not a big deal and might safely be ignored.

Table 3: The Determinants of the Number of Presidential Veto Overrides

Dependent Variable: Number of Presidential Veto Overrides			
Regressor	Poisson	Poisson (exposure/offset)	Poisson
MarginHouse	0.004	0.0003	0.002
	(0.005)	(0.005)	(0.005)
MarginSenate	-0.06***	-0.04*	-0.05**
	(0.02)	(0.02)	(0.02)
CongressExperience	1.90**	1.77**	1.82**
	(0.74)	(0.76)	(0.75)
GubernatorialExperience	1.97**	2.45***	2.28***
	(0.79)	(0.82)	(0.82)
Reelect	0.46	0.04	0.18
	(0.35)	(0.38)	(0.39)
PresidentVote	0.05*	0.01	0.02
	(0.03)	(0.03)	(0.03)
Exposure(NumberVetos)		1	
ln(PresidentVetos)			0.70**
			(0.29)
Constant	-4.33**	-4.84**	-4.52**
	(1.72)	(1.99)	(1.90)
Log likelihood	-35.41	-32.26	-31.72
Observations	26	25	25

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-tailed)
Standard errors are given in parentheses

3 Overdispersion, Underdispersion, and Event Count Models

3.1 What is Over- and Underdispersion?

Event count models are similar to our other binary and multichotomous models that we have examined in that they can be characterized as a realization of a latent process. In the case of event counts, the latent process is the *rate* at which events occur. For example, consider watching animals. Suppose we decide to count the number of cats that come to our backyard each day. Over two weeks, we might observe Y_{cats}={0, 1, 1, 0, 2, 0, 1, 0, 3, 1, 2, 1, 0, 2}. This yields $\bar{Y}_{cats}$=1 and $\sigma_{cats} = 0.92$. There is some underlying rate at which events occur such that we'd expect some number of cats to pass through on a particular day - let's call this λ_{cats}. Say we're interested in the probability that we observe 0, 1, 2, 3, ... cats per day. To figure out this probability we might make the following 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 might seem unrealistic, but in other cases it isn't).
3. The periods are all of the same length (this is not critical because we have seen we can deal with it using our exposure trick).
4. The probability of an event occurring (i) is constant within a particular period and (ii) is independent of other events during the same period (this assumption is critical).

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

We have:

$$\Pr(y_t = y) = \frac{e^{-\lambda}\lambda^y}{y!} \tag{24}$$

for $\lambda > 0$, $y \in \{0, 1, 2, \ldots\}$. The parameter λ is the unobserved rate of occurrence. It is also the expected value of the variable Y i.e. $E[Y] = \lambda$. If we use the exponential link function to allow the mean to vary according to some covariates, we have $\lambda_i = e^{X_i\beta}$. This yields the log-likelihood from before i.e. The log likelihood is:

$$\ln\mathcal{L} = \sum_{i=1}^{N}[-e^{X_i\beta} + y_iX_i\beta - \ln(y_i!)] \tag{25}$$

The last 'two-part' assumption we just made is critical to what the event count data generated will look like:

- Assumption of independence
- Assumption of constant rates

Let's examine these assumptions in turn.[4]

3.1.1 Independence and Contagion

Antelopes

Since antelopes are herd animals, when you see one, you'll likely see some more. This suggests that counts of antelopes will probably violate the assumption of independent events i.e. the assumption that one event has no effect on the likelihood of observing additional events in the same period. In the case of antelopes, one event (observation of an antelope) increases the likelihood of another - this is called **positive contagion**. If we have positive contagion, then we are likely to see greater numbers of higher and lower counts. For example, we might go nearly two weeks without seeing any antelope, but then see seven each on the last two days i.e. $Y_{antelope}$={0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 7}. As with our cat example, this count of antelopes has $\bar{Y}_{antelope}$=1. The difference is that $\sigma_{antelope} = 6.46$ rather than $\sigma_{cats} = 0.92$. In effect, positive contagion increases the variance of the observed counts. We call this **overdispersion**.

This causes problems for a Poisson model. Recall that in a Poisson model, $E[Y] = \text{Var}(Y) = \lambda$. In other words, the Poisson model imposes the assumption that the mean is the same as the variance. If we used the Poisson model when we had overdispersed data we'd effectively be requiring the variance to be less than it really is. As a result, we'd underestimate the true variability of the data, causing us to underestimate our standard errors, and hence to overestimate the precision of our coefficients etc.

Foxes

In some respects, foxes are the opposite of antelopes. Foxes are quite territorial and, as a result, seeing one fox means that it is unlikely that you will see another fox any time soon. Now, the occurrence of one event decreases the probability of another event in the same period - again, we are violating the assumption of independent events. Unsurprisingly, this is called **negative dispersion** and it yields greater numbers of counts right around the mean. This is known as **underdispersion**. As an example, we might see Y_{foxes}={1, 0, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1}. As with our cat and antelope examples, this count of foxes has $\bar{Y}_{foxes}$=1. The difference is that now $\sigma_{foxes} = 0.15$ instead of $\sigma_{antelope} = 6.46$ and $\sigma_{cats} = 0.92$. In effect, we have less variability than for an independent Poisson process. If we used the Poisson model when we had underdispersed data, we'd artificially overestimate the variability of Y_{foxes}, causing us to overestimate our standard errors, and hence to underestimate the precision of our coefficients etc.

Cross-Period Effects

Note that if our time periods are arbitrary, then contagion (positive or negative) *within* observations also implies contagion *across* observations. Like contagion within observations, contagion across observations can also lead to over- or underdispersed data relative to the Poisson process.

[4] Although what follows is based on notes from Zorn, you might also want to look at King (1989*b*).

3.1.2 Heterogeneity

The assumption of constant rates 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 – there is likely to be **heterogeneity**. For example, we might think that there is heterogeneity in the counts of presidential vetoes during a Congressional session. If a president is more likely to veto bills, say, early in a session (when 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. Unobserved heterogeneity of this type also leads to overdispersion, $E[Y] < \text{Var}(Y)$, of a form exactly the same as that for positive contagion outlined above. This is because the non-constant rate induces greater random variability in Y than would a constant λ.

3.1.3 Summing Up

Formally, the Poisson model requires that $E[Y_i] = \text{Var}(Y_i) = \lambda_i$. Let's imagine that we relax this mean-variance equality restriction and say that $\text{Var}(Y) = \lambda_i \sigma$. Then, in general, we can think of three situations:

1. Poisson Dispersion: $\leftrightarrow E[Y] = \text{Var}(Y) \leftrightarrow \sigma = 1$
2. Overdispersion: $\leftrightarrow E[Y] < \text{Var}(Y) \leftrightarrow \sigma > 1$
3. Underdispersion: $\leftrightarrow E[Y] > \text{Var}(Y) \leftrightarrow 0 < \sigma < 1$

On the whole, political science data are not Poisson distributed, even conditionally. We typically study dependent processes and/or we regularly have unobserved heterogeneity because we fail to measure things properly.

3.2 Dealing with Overdispersion

In political science, overdispersion is much more common than underdispersion. Moreover, in most instances where we have covariates x_i, what we really care about is conditional over- or underdispersion i.e. over- or underdispersion in the errors.

3.2.1 Testing for Overdispersion

One way to test for overdispersion is to 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 following equation:

$$\hat{z}_i = \delta \hat{\lambda}_i + \epsilon_i \tag{26}$$

where

$$\hat{z}_i = \frac{(y_i - \hat{\lambda}_i)^2 - y_i}{\hat{\lambda}_i \sqrt{2}} \tag{27}$$

and $\hat{\lambda}_i$ is the predicted value of λ_i for observation i (that is $e^{x_i \hat{\beta}}$) from the Poisson regression. So the test has three steps.

1. Estimate a poisson regression of y_i on x_i and generate predicted counts $\hat{\lambda}_i$
2. Calculate $\hat{z}_i$ according to Eq. (27) above.
3. Estimate Eq. (26) using OLS and test $H_0 : \hat{\delta} = 0$.

As we'll see in a moment, there are a bunch of other tests for overdispersion that are based on the Negative Binomial model.

3.2.2 Negative Binomial Model

We use a Negative Binomial (NB) model to deal with overdispersion. In the NB model, we still have the mean as $E[Y_i] = \lambda_i = e^{X_i \beta}$. However, we now assume that the variance is

$$\sigma^2 = \lambda_i(1 + \alpha\lambda_i) \tag{28}$$

How did we come up with this? The following is based on Long (1997, 230-238) and Zorn's notes.

In the Poisson model, we stated that λ_i was fully determined by a linear combination of the Xs. The negative binomial essentially adds some unobserved heterogeneity, such that λ_i is determined by the Xs and some unobservable, observation-specific random effect ϵ_i, which is assumed uncorrelated with the Xs. So we now have $\tilde{\lambda}_i$ such that

$$\begin{aligned} \tilde{\lambda}_i &= e^{x_i \beta + \epsilon_i} \\ &= e^{x_i \beta} e^{\epsilon_i} \\ &= \lambda_i \delta_i \end{aligned} \tag{29}$$

where $\delta_i = e^{\epsilon_i}$. The negative binomial is not identified without an assumption about the mean of the error term and the most convenient assumption is that $E[\delta_i] = 1$ since this gives us $E[\lambda_i] = \lambda_i$. This is good because it means that we have the same expected count as from the Poisson model.

Since the distribution of observations is still Poisson given X and δ, we have

$$\Pr(y_i | x_i, \delta_i) = \frac{e^{-\tilde{\lambda}_i} \tilde{\lambda}_i^{y_i}}{y_i!} = \frac{e^{-\lambda_i \delta_i} (\lambda_i \delta_i)^{y_i}}{y_i!} \tag{30}$$

Since δ_i is unknown, we cannot compute $\Pr(y_i | x_i, \delta_i)$ and instead need to compute the distribution of y given only X. To compute $\Pr(y_i | X_i)$ without conditioning on δ, we average $\Pr(y_i | X_i)$ by the probability of each value of δ. So, if g is the pdf of δ then

$$\Pr(y_i|x_i) = \int_0^\infty [\Pr(y_i|x_i, \delta_i) \times g(\delta_i)] d\delta_i \tag{31}$$

In order to be able to solve this equation, we need to specify the form of the pdf for δ. Most people assume that δ_i has a gamma distribution with parameter ν_i.

$$g(\delta_i) = \frac{\nu_i^{\nu_i}}{\Gamma(\nu_i)} \delta_i^{\nu_i - 1} e^{-\delta_i \nu_i} \text{ for } \nu_i > 0 \tag{32}$$

where $\Gamma(\nu_i) = \int_0^\infty t^{\nu-1} e^{-t} dt$. The nice thing about this distribution is that $E[\delta_i] = 1$ which was the convenient assumption we wanted to make since it means that the mean function of the negative binomial is the same as the poisson. $\text{Var}(\delta_i) = \frac{1}{\nu_i}$. Using (30) and (32) to solve (31), we obtain the following negative binomial (NB) probability distribution:

$$\Pr(y_i|x_i, \nu_i) = \frac{\Gamma(y_i + \nu_i)}{y_i! \Gamma(\nu_i)} \left(\frac{\nu_i}{\nu_i + \lambda_i} \right)^{\nu_i} \left(\frac{\lambda_i}{\nu_i + \lambda_i} \right)^{y_i} \tag{33}$$

The expected value of this is the same as for the Poisson distribution i.e. $E[y_i|X_i] = \lambda_i$. However, the conditional variance now differs:

$$\text{Var}(y_i|x_i) = \lambda_i \left(1 + \frac{\lambda_i}{\nu_i} \right) = e^{x_i \beta} \left(1 + \frac{e^{x_i \beta}}{\nu_i} \right) \tag{34}$$

Since λ and ν are both positive, the variance will exceed the conditional mean. This increases the relative frequency of low and high counts. So far, so good. But the variance remains unidentified since if ν_i varies by observation, we will have more parameters than observations. The typical assumption is that ν_i is the same for all observations i.e.

$$\nu_i = \alpha^{-1} \text{ for } \alpha > 0 \tag{35}$$

This particular formulation just helps simplify some equations. If we substitute (35) into (34), we have:

$$\text{Var}(y_i|x_i) = \lambda_i \left(1 + \frac{\lambda_i}{\alpha^{-1}} \right) = \lambda_i (1 + \alpha \lambda_i) = \lambda_i + \alpha \lambda_i^2 \text{ for } \alpha > 0 \tag{36}$$

The bottom line is that the NB Model has:[5]

- $E[Y] = \lambda_i = e^{X_i \beta}$
- $\text{Var}(Y) = \lambda_i (1 + \alpha \lambda_i), \ \ \alpha > 0$.
  - Thus, the variance is always greater than the mean, but ... the variance is still (positively) dependent on the mean i.e. heteroskedastic.

[5]It turns out that there is actually a second specification. The one outlined in the notes is sometimes referred to as Negative Binomial 2 or NB2 - it is the most common one. Negative Binomial 1 (NB1) has the same expected count as NB2 i.e. $E[Y] = \lambda_i = e^{X_i \beta}$. The difference comes in terms of the variance. In the NB1 model, the variance is $\text{Var}(Y) = \lambda_i + \alpha \lambda_i$ instead of $\lambda_i + \alpha \lambda_i^2$. You can get STATA to estimate either model, though, the default is the NB2 model that I have outlined in the notes.

- Larger values of α correspond to greater amounts of overdispersion.

- The density of the NB distribution is:

$$\Pr(y_i|x_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{y_i!\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i}\right)^{\alpha^{-1}} \left(\frac{\lambda_i}{\alpha^{-1} + \lambda_i}\right)^{y_i} \quad (37)$$

- While some people say that the Negative Binomial model corresponds to the Poisson model when $\alpha = 0$, this is not exactly true since the NB model is unidentified when $\alpha = 0$. However, if α is small or insignificant, we can go with the Poisson model. Another test for overdispersion in our data is to run an NB model and do a Wald test or an LR test for whether $\alpha = 0$.[6] If α is not significantly different from 0, we can just use a Poisson model (we don't need to use the NB model).

- Because α is restricted to be greater than 0, most statistical packages actually estimate $\frac{1}{\alpha}$ or $\ln\left(\frac{1}{\alpha}\right)$.

- Finally, the model remains log-linear and we still have $E[\hat{Y}_i] = e^{x_i\hat{\beta}}$. As a result, we can still interpret our results in terms of Incident Rate Ratios and in terms of predicted counts (and first differences). We could also calculate the predicted probabilities for discrete counts conditional on $x_i\hat{\beta}$ - though, we'd probably not want to do this last one by hand!

Note that we derived the negative binomial by thinking about unobserved heterogeneity across observations. However, we can also think about it as being derived from a process of positive contagion (see King (1989*b*)).

3.2.3 Estimating the Negative Binomial Model

The likelihood function for the NB model is:

$$\begin{aligned}\mathcal{L} = \prod_{i=1}^{N} \Pr(y_i|x_i) &= \prod_{i=1}^{N} \frac{\Gamma(y_i + \alpha^{-1})}{y_i!\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\nu_i + \lambda_i}\right)^{\alpha^{-1}} \left(\frac{\lambda_i}{\alpha^{-1} + \lambda_i}\right)^{y_i} \\ &= \prod_{i=1}^{N} \frac{\Gamma(y_i + \alpha^{-1})}{y_i!\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\nu_i + e^{x_i\beta}}\right)^{\alpha^{-1}} \left(\frac{e^{x_i\beta}}{\alpha^{-1} + e^{x_i\beta}}\right)^{y_i} \quad (38)\end{aligned}$$

After taking the log of the likelihood function, we would maximize it with respect to β and α. To estimate the NB model in STATA type:

```
nbreg nover  hmargin smargin congexpr govexpr reelect popvote
```

You can deal with any exposure issues as before. Results from this model are shown in Table 4 above.

[6]Note that the tests are not your usual Wald or LR test in this case. This is because that fact that α must be greater than or equal to 0, the asymptotic distribution of $\hat{\alpha}$ is only half of a normal distribution i.e. all values less than 0 have a probability of 0. This requires an adjustment to the usual significance level of the test. STATA takes this into account when it provides the results from an LR test at the bottom of its output. I make this point because if you conducted your own LR test by typing 'lrtest' after the estimation command, this would give you the wrong output. For a good discussion of this issue, see Long and Freese (2006).

Table 4: The Determinants of the Number of Presidential Veto Overrides

Dependent Variable: Number of Presidential Veto Overrides				
Regressor	Poisson	Poisson (exposure)	Poisson	Negative Binomial
MarginHouse	0.004	0.0003	0.002	0.004
	(0.005)	(0.005)	(0.005)	(0.005)
MarginSenate	-0.06***	-0.04*	-0.05**	-0.06***
	(0.02)	(0.02)	(0.02)	(0.02)
CongressExperience	1.90**	1.77**	1.82**	1.90**
	(0.74)	(0.76)	(0.75)	(0.74)
GubernatorialExperience	1.97**	2.45***	2.28***	1.97**
	(0.79)	(0.82)	(0.82)	(0.79)
Reelect	0.46	0.04	0.18	0.46
	(0.35)	(0.38)	(0.39)	(0.35)
PresidentVote	0.05*	0.01	0.02	0.05*
	(0.03)	(0.03)	(0.03)	(0.03)
Exposure(NumberVetos)		1		
ln(PresidentVetos)			0.70**	
			(0.29)	
Constant	-4.33**	-4.84**	-4.52**	-4.33**
	(1.72)	(1.99)	(1.90)	(1.72)
α				1.20e-7
				(0.0002)
Log likelihood	-35.41	-32.26	-31.72	-35.41
Observations	26	25	25	26

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-tailed)
Standard errors are given in parentheses.

3.3 Dealing with Underdispersion

A similar distribution to the NB distribution – the 'continuous parameter binomial' (CPB) is used to model underdispersed data i.e. data which has negative contagion. This is essentially a variant of the binomial distribution that is scaled to ensure that the probabilities sum to 1. The density for the CPB model is:

$$P(y_i|X_i,\alpha) = \frac{\frac{\Gamma\left(\frac{-\lambda_i}{\alpha-1}+1\right)}{y_i!\Gamma\left(\frac{-\lambda_i}{\alpha-1}-y_i+1\right)}(1-\alpha)^{y_i}(\alpha)^{\frac{-\lambda_i}{\alpha-1}-y_i}}{D_i} \tag{39}$$

where D_i is the aforementioned scaling factor which happens to be the sum of 0 to $\frac{\lambda_i}{\alpha-1}+1$ of the binomial distribution. This model has the following characteristics:

- As before, $E[Y_i] = \lambda_i$, where we typically have $\lambda_i = e^{X_i\beta}$.
- Var$(Y) = \lambda_i\alpha$ with $0 < \alpha < 1$
- This model essentially becomes the Poisson when $\alpha = 1$.

- The CPB model imposes a theoretical 'upper limit' on the count variable. In particular, $\max(Y_i) = \frac{\lambda_i}{\alpha-1}$. This limit is due to the fact that the variability of Y is constrained by α. In effect, initial events reduce the probability of future events (due to negative contagion) and, therefore, also reduce the maximum number of events that could occur in a period.

The log-likelihood of the CPB model is:

$$\ln\mathcal{L}_{CPB} = \sum_{i=1}^{N} \left\{ \ln\Gamma\left(\frac{\lambda_i}{\alpha-1}+1\right) - \ln\Gamma\left(\frac{-\lambda_i}{\alpha-1} - y_i + 1\right) + y_i \ln(1-\alpha) + \left(\frac{\lambda_i}{\alpha-1} - y_i\right)\ln(\alpha) - \ln(D_i) \right\} \tag{40}$$

It turns out that in practice, underdispersed data is quite rare. STATA does not estimate a CPB model. King (1989*b*, 771) says that he is unaware of anyone using the CPB model in the social sciences.

3.4 A Generalized Event Count (GEC) Model

King (1989*b*) provides a 'generalized event count' distribution – a single equation model that can take on any of the three possible types of dispersion: overdispersion, underdispersion, or mean-variance equality. The details of this model can be found in King (1989*b*).

- Like the NB and CPB models, the GEC has a single parameter that indexes the ratio of the variance to the mean. King refers to this paramter as σ^2.
  - When $\sigma^2 = 1$, we essentially get the Poisson model.
  - When $\sigma^2 > 1$, we essentially get the NB model.
  - When $0 < \sigma^2 < 1$, we essentially get the CPB model.
- Because the GEC model (sort of) nests the CPB model, it too has a restriction on the maximum number of counts.
- You can use Gary King's program COUNT to estimate GEC and other event count models. Go to his webpage for this.

3.5 Modeling Variance in Event Counts

Just as in the regression context, as well as binary and ordered dependent variables, it is possible to allow the variance of an event count to vary systematically with a set of covariates - let's call them z_i. This is typically done in the context of the NB model by setting $\sigma_i = e^{z_i\gamma}$. This yields an NB model in which the variance term α depends explicitly on z_i. The likelihood of this model is the same as in Eq. 38 except that $\sigma^2 \equiv \frac{1}{\alpha} = e^{z_i\gamma}$. After taking the log, we now maximize this with respect to β and γ. Other than that, interpretation is exactly the same. To estimate this model in STATA you use the GNBREG command. We would use this type of model when we thought that the variance of the count (matters related to contagion or heterogeneity) is dependent on observable, systematic factors. King (1989*a*) provides an applied example of this model.

3.6 An Important Aside

When talking about over- and underdispersion, we usually mean conditional on the effects of the covariates i.e. Var$(y_i|x_i\beta)$. When we introduce covariates, we are factoring out some of the heterogeneity in the data – it is no longer unobserved. This means that as the model gets better, we often see our data go from overdispersed, to Poisson, to underdispersed. This suggests that we should be very careful about drawing substantive inferences about overdispersion etc. in our models. In effect, matters of dispersion are related to our models and are not an abstract property of the world - we can get rid of overdispersion etc. by putting in the right variables. I made a similar point last week about the dangers of drawing substantive inferences about IIA. I'll make a similar point about drawing inferences about time dependence from our duration models in a week or two.

4 Truncated and Censored Counts

5 Truncation

Sometimes because of the way our data are collected, we only observe $y_i \geq 1$. For example, imagine that we want to know how many times people ride a bus each week. One way to gather the required data is to sit on a bus and ask people how often they ride the bus each week. This is a good way to get information on the decision of whether to ride at all, but we get no information on the large number of 'no-rides' i.e. people who never ride the bus - we just never meet those people. In effect, our data is truncated at 0. If we use a Poisson or NB model, we are falsely allowing the probability of zero to be positive. The way to deal with this problem is to fit what is known as a **zero-truncated Poisson** or **NB** model. Let's look at the zero-truncated Poisson model to see how this works. You start with the basic Poisson model.

$$\text{Pr}(y_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \tag{41}$$

For a given set of Xs, the probability of observing a zero is Pr$(y_i = 0) = e^{-\lambda_i}$. Thus, the probability of observing a nonzero count is just Pr$(y > 0) = 1 - P(y_i = 0) = 1 - e^{-\lambda_i}$. By the law of conditional probability, Pr$(A|B) = \frac{\text{Pr(A and B)}}{\text{Pr}(B)}$. Thus, the probability of observing a specific value of y given that we know the count is not zero is:

$$\begin{aligned}\text{Pr}(y_i|y_i > 0) = \frac{\text{Pr}(y_i)}{\text{Pr}(y_i > 0)} &= \frac{\text{Pr}(y_i)}{1 - e^{-\lambda_i}} \\[2em] &= \frac{\left(\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}\right)}{1 - e^{-\lambda_i}}\end{aligned} \tag{42}$$

In effect, we are simply increasing each unconditional probability by the factor $\{1 - e^{-\lambda_i}\}^{-1}$ so that the probability mass of the truncated distribution sums to 1.

There are two types of expected count that you might be interested in with zero-truncated models: unconditional count and the conditional count. First, you might be interested in the expected number of bus rides in the entire population, not just among those who you know ride the bus at least once. In the zero-truncated Poisson model, this unconditional count is:

$$E[y_i] = \lambda_i = e^{x_i\beta} \tag{43}$$

Second, you might be interested in the expected number of bus rides among those who ride the bus at least once. In the zero-truncated Poisson model, this conditional count is

$$E[y_i|y_i > 0] = \frac{\lambda_i}{1 - e^{-\lambda_i}} \tag{44}$$

For the zero-truncated negative binomial model, the unconditional expected count is:

$$E[y_i] = \lambda_i = e^{x_i\beta} \tag{45}$$

and the conditional expected count is:

$$E[y_i|y_i > 0] = \frac{\lambda_i}{1 - (1 + \alpha\lambda_i)^{-\frac{1}{\alpha}}} \tag{46}$$

To estimate a zero-truncated model in STATA, you use the ZTP or the ZTNB commands.

6 Censored

Sometimes because of the way our data are collected, the highest event count is recorded as N or more i.e. you might have a few cases with a very large number of events, but you record them as N or more. To deal with this problem, you just have to change the last entry in the likelihood function to $\Pr(y_i \geq N)$ rather than $\Pr(y_i = N)$. This is relatively easy to do.

7 Hurdle Models and Zero-Inflated Count Models

Note that the solution outlined above to the issue of data that is truncated at zero was to assume that the process of going from a count of zero to a count of one was the same Poisson (or NB) process as going from one to two etc. In effect, we assumed that the data all follow a Poisson (or NB) process but that we just decided not to observe (or record) the zeros. This is a very strong assumption to make! It might make more sense to assume that the process which governs zeros and non-zeros is different from the process which determines how many times something is done given that something is done. Obviously you need information on the zeros and non-zeros to estimate this type of model. The point here is that if we think that the process determining whether an event at all takes place is different from the process determining how many events occur given that one occurs, then the various event count models that I have just outlined are not appropriate. Instead, we would want to use a hurdle model or zero-inflated model. These models are also appropriate for estimating event count models in which the data generating process results in a larger number of zero counts than would be expected under the standard assumptions of the Poisson or NB distributions. As Zorn (1998) notes, hurdle models and zero-inflated models are both specific forms of a more general 'dual regime' model.

What's the basic difference between hurdle models and zero-inflated models? Basically, in a hurdle model, we assume that there are two types of people: (i) those who never experience an event and (ii) those who always experience at least one event. In contrast, in a zero-inflated model, we assume that there are two types of people: (i) those who can never experience an event and (ii) those who might experience an event but don't have to. As you can see, the difference between these models is in terms of how we conceptualize the second type of person in each case. Below, I go into more detail.

7.1 Hurdle Models

A hurdle model combines a binary model to predict zeros and a zero-truncated Poisson (or NB) model to predict non-zero counts. Let y be a count outcome that ranges from zero to some maximum value. Suppose that zero counts are generated by a binary process. We'll use a logit model to model the binary outcome $y = 0$ versus $y > 0$:

$$\text{Pr}(y_i = 0) = \frac{e^{x_i\beta}}{1 + e^{x_i\beta}} = \pi_i \tag{47}$$

Positive counts are assumed to be generated by either a zero-truncated Poisson (or NB) process. In effect, there are two equations where zero is viewed as a 'hurdle' that you have to get past before reaching positive counts. Since we have two equations, we can have different covariates in the different equations. The predicted rates and probabilities from the hurdle model are computed by mixing the results from the binary model and the zero-truncated model. The probability of a zero is:

$$\text{Pr}(y_i = 0) = \pi_i \tag{48}$$

Since positive counts can only occur if you get past the zero hurdle (which occurs with probability 1-π_i), we weight the conditional probability from the zero-truncated model i.e.

$$\text{Pr}(y_i = 1) = (1 - \pi_i)\text{Pr}(y_i|y_i > 0) \text{ for } y > 0 \tag{49}$$

Thus, the unconditional rate is computed by combining the mean rates for those with $y = 0$ (which is 0) and the mean rate for those with positive counts i.e.

$$\begin{aligned} E[y_i] &= \pi_i \times 0 + \{(1 - \pi_i) \times E(y_i|y_i > 0)\} \\ &= (1 - \pi_i) \times E(y_i|y_i > 0) \end{aligned} \tag{50}$$

where $E(y_i|y_i > 0)$ is defined by Eq. (44) for the zero-truncated poisson model and by Eq. (46) for the zero-truncated negative binomial model.

7.2 Estimating a Hurdle Model

STATA does not have an automatic command for the hurdle model. However, it is relatively easy for you to get STATA to give you the results that you want.[7]

First, generate a variable equal to one if the count variable equals zero.

```
gen zero=(num==0);
label var zero "1 = No count, 0 = positive count";
```

Then estimate a binary dependent variable model (logit or probit) to explain the observation of a zero.

```
probit zero X;
```

Remember that since a zero outcome is the prediction, the signs of the coefficients have precisely the opposite interpretation from that of a normal logit or probit. In other words, a positive coefficient indicates that the variable increases the probability of zero events and a negative coefficient indicates that the variable decreases the probability of zero events. Then estimate a zero truncated count model for the rest of the data.

```
ztnb num X if zero==0;
```

You can then present the two sets of results in two columns of a table.

7.3 Interpreting a Hurdle Model

The interpretation of hurdle models can be somewhat difficult, particularly if you have variables in both parts of the models. One suggestion is to generate predicted probabilities or graphs of predicted probabilities for various event counts. This can be done using the PRVALUE command from Long's SPOST ado file (you can, of course, also do it manually).

```
logit zero X;

est store hlogit;

ztnb number X if zero==0;

est store hztnb;

est restore hlogit;
```

[7] For more information, see Long and Freese (2006, 387-393).

```
prvalue, x(X=some value) rest(mean);
```

PRVALUE saves its results in the matrix PEPRED. You can see this matrix by typing:

```
matrix list pepred;
```

We need the information in the second row and second column on this matrix i.e. the probability of a zero count. We can save this in a scalar variable for later.

```
scalar hrdlp0=pepred[2,2];
```

Now we need to restore the zero-truncated part of the model and get the predicted probability from that;

```
est restore hztnb;

prvalue, x(X=some value) rest(mean) all maxcnt(19);
```

Note that when using PRVALUE now, you need to specify the maximum count for which you want to calculate a predicted probability (MAXCNT). You also need to include the ALL term to ensure that all of the observations in the sample are used and not just those in the estimation sample. This is important because the ZNTB model was estimated on just those observations with positive counts, but we want the means to be based on the sample used for for the LOGIT model. PRVALUE saves its results in two matrices: PEPRED and PECPRED. In this case, we want the conditional predictions and so we need to use PECPRED.

```
matrix list peCpred;
```

We now want to calculate the predicted probabilities for the positive counts. We do this by looping through the count values 1-19, retrieving the required quantities from the PECPRED matrix, and applying the formula shown in Eq. (49). We first generate a variable PPROBDIST that stores the predicted probabilities.

```
gen Pprobdist=.;

gen count=_n-1;

forvalues i=1 (1) 19 {;
    scalar ztnbp`i'=peCpred[2, `i'+1];
    scalar hrdlp`i'=(1-hrdlp0)*ztnbp`i';
    replace Probdist = hrdlp`i' if count==`i';
    display "Prob(article=`i'|X) = " hrdlp`i';
    };
```

```
display "Prob(article=`i'|X) = " hrdlp0;

replace Pprobdist=hrdlp0 if count==0;
```

You can then graph the predicted probabilities for the different counts by typing:

```
graph twoway bar Pprobdist count if count<=19, barwidth(1);
```

You can, of course, repeat this process, once each for different values of the variables of interest. In each process, store the predicted probabilities under different variable names (i.e. PPROBDIST1 and PPROBDIST2). Two conditional predicted probability distributions can then be overlaid on one another to demonstrate the effect of a change in a variable of interest.

In addition to calculating the predicted probabilities for various event counts, you can also calculate the expected number of events for different scenarios using Eq. (50).

Note that with the hurdle model, you are not permitted to get zero counts once the 'hurdle' has been crossed. As I noted earlier, this distinguishes the hurdle model from the zero-inflated model that I am about to present. To see an example of a hurdle model, see King (1989*a*).

7.4 Zero-Inflated Models

In terms of the zero-inflated models, we can think that there are two groups of people. The first group consists of people who always have zero counts. The second group consists of people who may have positive counts but may for some reason have a zero count. In effect, you have a group of people who have counts of zero because they didn't do something but might have, and a group of people who would never have even considered doing the thing. Although there are two equations in the hurdle model, there is assumed to be only one process by which a zero can be produced. In the zero-inflated model, we assume that there are two different processes that might produce a zero.

The zero-inflated model assumes that there are two latent (unobserved) groups. An individual in the Always Zero group (Group A) has an outcome zero with a probability of 1, whereas an individual in the Not Always Zero group (B) might have a zero count. This process has three steps: (i) model membership into the latent groups, (ii) model counts for those in group B, (iii) compute observed probabilities as a mixture of the probabilities for the two groups.

Step 1: Membership in Group A

Let $A = 1$ if someone is in group A, 0 otherwise. We can use a logit or probit to model group membership.

$$\psi_i = \text{Pr}(A_i = 1) = F(x_i\gamma) \tag{51}$$

where ψ_i is the probability of being in Group A for individual i. The z-variables are known as inflation variables because they inflate the number of zeros etc.

Step 2: Counts for those in Group B

Among those who are not always zero, the probability of each count (including zeros) is determined either by the Poisson or NB distribution. Note that the probability equation below is conditioned on the xs and on $A_i = 0$. For the zero-inflated Poisson (ZIP) model, we have

$$\Pr(y_i|x_i, A_i = 0) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \tag{52}$$

Step 3: Mixing Groups A and B

We now need to mix the two groups according to their proportions in the sample to determine the overall rate. The proportion in each group is defined as:

$$\begin{aligned} \Pr(A_i = 1) &= \psi_i \\ \Pr(A_i = 0) &= 1 - \psi_i \end{aligned}$$

The probabilities of a zero within each group are

$$\begin{aligned} \Pr(y_i = 0|A_i = 1, x_i, z_i) &= 1 \text{ by definition} \\ \Pr(y_i = 0|A_i = 0, x_i, z_i) &= \text{outcome of the Poisson or NB model} \end{aligned} \tag{53}$$

Thus, the overall probability of a zero count is

$$\begin{aligned} \Pr(y_i = 0|x_i, z_i) &= [\psi_i \times 1] + \{(1 - \psi_i) \times \Pr(y_i = 0|x_i, A_i = 0)\} \\ &= \psi_i + \{(1 - \psi_i) \times \Pr(y_i = 0|x_i, A_i = 0)\} \end{aligned} \tag{54}$$

For outcomes other than zero, we have

$$\begin{aligned} \Pr(y_i = k|x_i, Z_i) &= [\psi_i \times 0] + \{(1 - \psi_i) \times \Pr(y_i = k|x_i, A_i = 0)\} \\ &= (1 - \psi_i) \times \Pr(y_i = k|x_i, A_i = 0) \end{aligned} \tag{55}$$

Expected counts are computed similarly

$$\begin{aligned} E(y_i|X_i, Z_i) &= [0 \times \psi_i] + \{\lambda_i \times (1 - \psi_i)\} \\ &= \lambda_i(1 - \psi_i) \end{aligned} \tag{56}$$

We esimate the zero-inflated Poisson and NB models using the STATA commands ZIP and ZINB.

STATA code to estimate these models is:

- zip dependentvariable independentvariables, inflate(independentvariables)
- zinb dependentvariable independentvariables, inflate(independentvariables)

For more information on these models, including how to interpret results and tests of model specification, see Long (1997, 242-247) and Long (2006, 398-405).

8 Temporal Dependence in Event Count Models

Notes to come, but see Brandt et al. (2000) and Brandt and Williams (2001).

References

Brandt, Patrick, Benjamin O. Fordham, Brian Pollins & John T. Williams. 2000. "Dynamic Modeling for Persistent Event Count Time Series." *American Journal of Political Science* 44:823–843.

Brandt, Patrick & John T. Williams. 2001. "A Linear Poisson Autoregressive Model: The Poisson AR(p) Model." *Political Analysis* 9:164–184.

King, Gary. 1988. "Statistical Models for Political Science Event Counts: Bias in Conventional Procedures and Evidence for the Exponential Poisson Regression Model." *American Journal of Political Science* 32:838–863.

King, Gary. 1989*a*. "Event Count Models for International Relations: Generalizations and Applications." *International Studies Quarterly* 33:123–147.

King, Gary. 1989*b*. "Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator." *American Journal of Political Science* 33:762–784.

King, Gary. 2001. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. New York: Cambridge University Press.

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. London: Sage Publications.

Long, J. Scott. 2006. *Regression Models for Categorical and Limited Dependent Variables Using STATA. Second Edition.* Texas: STATA Corporation.

Zorn, Christopher J. W. 1998. "An Analytical and Empirical Examination of Zero-Inflated and Hurdle Poisson Specifications." *Sociological Methods and Research* 26:368–400.