CHAPTER 32

Count Models in Criminology

JOHN M. MACDONALD AND PAMELA K. LATTIMORE

INTRODUCTION

Crime can be measured in many metrics, including occurrence (yes/no), seriousness (e.g., felony/misdemeanor), and frequency or rate. Frequency can be considered in two analogous ways – how soon until an occurrence (the rate) or how many occurrences per unit such as weeks, months, or years (counts). We can also look at frequency with respect to other units such as population groups or areal units like neighborhoods or counties.

Count models, including the basic Poisson regression model and its various approximations, are now extremely popular in criminology. For example, Gardner et al. (1995) showed the use of count models for predicting violent incidents in the community involving 797 individuals who were evaluated in an emergency room of a psychiatric hospital and tracked in the community for 6 months. Sampson and Laub (1996) relied on Poisson regression and its variant of negative binomial regression for assessing the effects of military experience on subsequent offending. Paternoster et al. (1997) used Poisson regression formulations to assess the effects of perceived fairness in arrest interactions for domestic violence on recidivism. Lattimore et al. (2004) estimated two types of Poisson regression models to examine the relationship of covariates on the average count of re-arrest and its variance on released California Youth Authority parolees. Osgood and Chambers (2000) examined county-level factors associated with counts for juvenile arrests for FBI index offenses using Poisson formulations. Parker (2004) similarly relied on count models to estimate changes in racially disaggregated counts of homicides in cities. Braga (2003) relied on the negative binomial version of the Poisson regression model to estimate changes in the counts of arrests among youth gun offenders in Boston over time. Lattimore et al. (2005) looked at the effects of substance abuse treatment on the count of rearrests among drug-involved probationers in Florida.

More generally, the importance of modeling criminal behaviors as counts has long been recognized (Blumstein et al. 1986; Greenberg 1991) and several decades of research in criminology has been devoted to studying the rate of crime among offenders according to a criminal career paradigm (see Piquero et al. 2003 for a review). Scholars of this perspective often refer to their interest in understanding Lambda (λ), the Greek symbol used in statistics to refer to an expected count or rate from a Poisson distribution. In some statistical texts, Lambda (λ) is simplified to the standard mean notation (denoted by the Greek symbol mu, μ) (see

A.R. Piquero and D. Weisburd (eds.), *Handbook of Quantitative Criminology*,
DOI 10.1007/978-0-387-77650-7_32,

Agresti 2007). We will be consistent with notation from statistics and will use the standard mean notation conditional on the Greek symbol λ to denote the expected average count or rate of crimes from a Poisson distribution.

Rates are an expression of counts according to some defined denominator (e.g., time or population) and, as such, the two are completely interchangeable. A number of approaches have been developed over the years for estimating counts that can be usefully applied to criminology. Today, statistical software programs allow us to compute most of the elegant solutions from statistics with ease. Caution is advised, however, and before using these methods, it is useful to understand the basic assumptions of each method. To this end, we will explain both the basic notation and methods for estimating regression models on count data and then provide a case study example from data collected on two samples of paroled offenders from the California Youth Authority in the 1980s.

WHY COUNT MODELS?

The nature of criminal behavior is such that for individuals or larger aggregations like neighborhoods, cities, or counties, the counts we observe are often small (even zero) for many units and occasionally large for a few units. Some of these findings are due to only observing these events when they are reported either in official crime records or self reports; however, many crimes like homicide are simply rare. In either case, the result is that when the average counts are small (e.g., less than 5), the distribution of outcomes is skewed to the right. A skewed distribution is problematic for ordinary least squares (OLS) regression analysis that assumes normal distribution in errors around the expected average $(E(Y = \mu))$. In this case, a method that relies on another distribution may be more appropriate. The Poisson distribution is unimodal and skewed to the right taking on values 0, 1, 2 .... The Poisson distribution is represented by a single parameter $\lambda > 0$, so its mean and variance are identical $(E(Y) = \mathrm{Var}(Y) = \lambda)$ (see Agresti 2007). The skewed nature of the Poisson distribution makes it an attractive probability distribution when one wants to estimate skewed crime counts.

It is important to note that as the number of counts gets larger, the distribution becomes more normally distributed (Agresti 2007). In other words, OLS is fine with count outcomes when the mean gets large (e.g., $\geq$ 20). So, although currently there seems to be some confusion in criminology journals that count data require a Poisson distribution, when the average counts are large, all else equal, it is better to estimate crime outcomes using OLS because the squared standard deviations from the mean are always the minimum estimator. As Angrist (2006) notes "You cannot improve on perfection" (p. 35). However, in many studies, the observed crime counts have an average low count of incidents and a skewed outcome distribution that presents real challenges for OLS estimations.

Statisticians, over a hundred years ago, recognized the problem when attempting to describe the probability of rare event counts and developed the Poisson probability distribution as an extension of a series of Bernoulli trials or binary events with probabilities of an event equal to p_i or no event equal to $(1 - p_i)$, where each event is assumed to occur independently of the next. As Good (1986, p. 160) notes, "Poisson's Law of Large Numbers stated that in a long sequence of trials the fraction of successes will very probably be close to the average of the "chances" for the individual trails even when the average does not tend to limit." There are a number of early examples of statisticians using the Poisson distribution, including the examination of Prussian soldiers that were kicked to death by horses (see Stigler 1986). For criminologists, the event of interest is often a crime, an arrest, a conviction, etc.

The basic formulation for the Poisson distribution is:

$$\Pr(Y|\lambda) = \frac{e^{-\lambda}\lambda^{Y}}{Y!} \quad \text{for } Y = 0, 1, 2, 3 \ldots \tag{32.1}$$

In (32.1), Y is the outcome represented by a count (e.g., count of crimes) and λ is a parameter representing the expected probability of the count according to a Poisson distribution.

For criminologists, one wants to know what predicts the number of crime counts assuming the probability of outcomes occurs via a Poisson distribution (Berk and MacDonald 2008). For example, in a simple experiment comparing the probability of committing future crimes for those randomly assigned to some treatment, one could compare the probability of crime counts for the treatment versus control condition. If arrests are a relatively rare outcome event that occurs independently of each other, then the Poisson distribution may be well suited.

However, criminology is typically focused on observational studies where there are a number of predictors and few well-controlled experiments. A solution to this situation was proposed by econometricians who expanded the Poisson probability distribution into the set of generalized linear models (GLM). These GLM are particularly useful for criminology when we are (1) trying to explain the expected response variable as counts ($E(Y = \text{crime counts}|\lambda)$) by a set of independent variables (X') and (2) for point of convenience and parsimony want to assume a linear relationship between outcomes and predictors or $E(Y|\lambda = x'\beta)$ (Greene 1997). Linear relationships are easier to visualize and provide a simplified method for explaining the expected distribution of crime counts at various units (e.g., individuals or places) and may provide an appropriate first-order approximation in the complex world of human behavior. The GLM transforms (32.1) to show that conditional on a set of independent variables (X') and a parameter μ that captures the distribution of observed outcomes (Y) that remains Poisson (Cameron and Trivedi 1998), as shown in (32.2).

$$\Pr(Y|X', \mu) = \frac{e^{-\mu}\mu^{Y}}{Y!} \quad \text{for } Y = 0, 1, 2, 3 \ldots \tag{32.2}$$

Readers interested in a more thorough coverage of GLM should consult approachable texts like Cameron and Trivedi (1998), Agresti (2007), and Long (1997).

In the remaining sections of this chapter, we will focus on the main GLM Poisson regression model and the regression variants of the negative binomial (NB) and zero-inflated Poisson (ZIP).

BASIC ASSUMPTIONS OF COUNT MODELS

Before demonstrating some applications of count models to criminology, we first discuss the basic assumptions of each approach. The Poisson regression and its variants are a form of regression analysis that models the expected rate, $E(\lambda)$ of observed crime outcomes (Y) according to a Poisson distribution. As discussed earlier, the Poisson probability distribution is skewed to the right. As a result, the logarithm of the Poisson probability distribution should be approximately linear. One can see this graphically in Fig. 32.1, where a distribution of observations is highly skewed to the right and the logarithmic smoother fits an almost straight line to the distribution. To get the expected count conditioned on $\lambda(E(\mu|\lambda))$ into linear form,

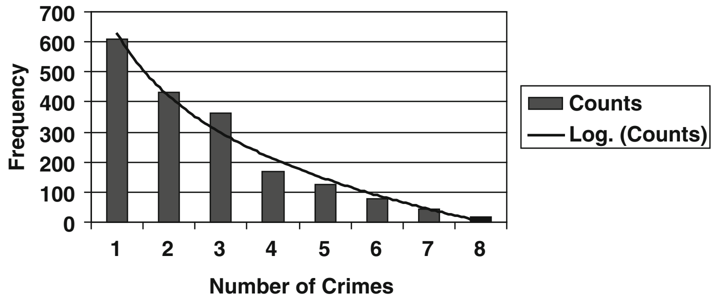

FIGURE 32.1. Distribution of crime counts.

Poisson regression relies on the canonical log-link function that provides the approximation between the linear predictors variables (denoted by X_k) and the mean of the expected distribution of counts (Cameron and Trivedi 1998). In other words, the Poisson regression model is a member of a class of log-linear models (Agresti 2007).

POISSON REGRESSION

The Poisson regression model can be expressed as the logarithm of the expected count outcome according to the following form:

$$\log(E(\mu|\lambda)) = \alpha + x'\beta \tag{32.3}$$

In (32.3), the expected average count of crime, μ, conditioned on λ is denoted by $(E(\mu|\lambda))$, which is a function of the intercept (α) plus a set of linear predictor variables (X'). Unlike in the OLS regression model that assumes predicted values follow a normal distribution defined by an expected average (μ) and a variance (σ^2), the Poisson distribution is defined by a single parameter, μ, where $E(Y) = \mu$ and var $(Y) = \mu$. Note for the Poisson regression this refers to the expected mean for outcomes and variance from the model – not the unconditional mean of the outcome and its variance. A number of criminology related articles express this assumption by simply examining the mean of the outcome and its variance. What one wants to see is whether the estimated mean from the regression model equals the variance.

The assumption that the conditional mean and variance are equal is rarely met with observational data in criminology. Instead, the variance is either less than or greater than the mean. When the conditional variance is less than the conditional mean, this is referred to as underdispersion. When the conditional variance is greater than the conditional mean, this is referred to as overdispersion. These terms simply mean that the model fit to the observed data has less or more variation than is expected according to the Poisson model. Underdispersion is often observed when estimating extremely rare counts of outcomes (e.g., arrests for murder in which the observed average of the outcome is zero and there are few observations with high counts). Overdispersion often occurs in count data with greater numbers of observations with more than fifty counts. One estimation solution for under or overdispersion is to correct by including a variance parameter in the estimating equation. Most statistical software packages that estimate Poisson regression set the variance parameter equal to 1 so that the conditional mean and variance are equal by default. These packages also generally provide tests of this assumption that the variance parameter equals 1. We discuss two examples of dispersion corrections – the negative binomial and zero-inflated Poisson – below.

NEGATIVE BINOMIAL REGRESSION

The negative binomial is the most commonly used correction for overdispersion. Osgood (2000) provides an excellent review of the negative binomial model and its application to aggregate crime counts, while Berk and MacDonald (2008) review the necessary assumptions for the negative binomial and how these assumptions may not be plausible in many circumstances. The basic formulation for the negative binomial is discussed in a number of statistical texts and posits that at different values of the observed outcome (Y = crime counts) one has a mixture of Poisson and gamma distributed data. The resulting distribution is one whose expected average outcome follows a Poisson distribution ($\lambda = \mu$) but whose variance follows a gamma distribution (var $\lambda = \mu^2/k$). A subset or marginal of a gamma mixture of Poisson distributions yields the negative binomial distribution (Agresti 2007). The negative binomial distribution, therefore, incorporates both distributions such that the expected mean of the outcome of counts (Y) follows a Poisson distribution and the variance is equal to the Poisson and gamma distribution, as shown in (32.4).

$$E(Y) = \mu, \quad \text{var}\,(Y) = \mu + \mu^2/k^{-1} \tag{32.4}$$

The parameter $k > 0$ describes the shape of the gamma distribution which is skewed to the right. It, therefore, follows that when $k = 0$, one returns to the original Poisson formulation because the expected mean and variance will be equal. The greater that k gets from zero, the greater the amount of overdispersion. For example, if the expected mean $\mu = 4$ and $k = 3$, the expected variance will equal 52. When $k < 0$, this is evidence of underdispersion because the expected variance will be less than the expected mean (e.g., expected mean $\mu = 4$ and $k = -3$ the expected variance will equal -44).

The basic extension of the Poisson Generalized Linear Model (GLM) to test for over- or under-dispersion is accomplished by adding a variance parameter (denoted by D^2)

$$\log(E(\mu|\lambda)) = \alpha + x'\beta + D^2 \tag{32.5}$$

When one rejects the null hypothesis that D^2 equals 1, this suggests that there is either under- or over-dispersion (see Agresti 2007). When D^2 equals 1 the model remains the standard Poisson GLM.

ZERO INFLATED POISSON REGRESSION

A potential problem with the standard Poisson regression analysis framework is that the counts may be extremely low such that most observations have values of zero and a few observations to the right-hand-side of the distribution drive the parameter estimates. As a result, one may have less variation (under-dispersion) than is captured in the expected mean when the zeros exceed what would be expected from a standard Poisson distribution. One way to deal with low counts within the Poisson framework is the zero-inflated Poisson (ZIP) model. The ZIP model works by combining the likelihood functions from the probit or logit and the Poisson regression models such that:

$$\log E(\mu|\lambda) = (1-\omega)^{*} \alpha + x'\beta \qquad (32.6)$$

Where ω captures the influence of covariates from extra zeros according to a logit model $(E(\omega) = \log(p_i/(1-p_i) = x'\beta)$. These models are less common in criminology, because criminology seems to more often confront cases where the counts on the right-end of an observed distribution exceed what one would expect from a simple Poisson process, hence there is greater variability than expected. However, as criminologists increasingly examine rare event outcomes, ZIP and other related models will increasingly be used.

A NOTE OF CAUTION

Before one abandons standard Poisson regression models and turns to other variants, we would like to caution readers against assuming that rejection of the null hypothesis that the variance parameter equals 1 provides prima fascia evidence that the Poisson model estimated is wrong. Having less or more variation than the expected average of the counts only implies that the Poisson regression model is the incorrect model if one assumes that the estimated model includes the correct predictors and the correct specification of the functional form. Alternatively, the "excess" variation may simply indicate that one has omitted an important variable or failed to incorporate relevant interactions between predictor variables that capture important heterogeneity between observations. Relying on theory and close scrutiny of the data in developing the model is essential, correcting the disturbance term is less important. After all, the model should describe the process that generated the data, rather than the data describing the process that generated the model. We refer readers to an article by Berk and MacDonald (2008) that has several illustrations of these points.

EXPRESSING COUNTS AND CALCULATING RATES

The Poisson regression model relies on the log-link function to provide a simple expression of the linear relationship between the predictor variables on the right hand side of (32.1) and the expected count. The expected count is obtained by exponentiating both sides of the equation. By taking the exponent of both sides of the equation, one obtains the expected average count, as demonstrated in (32.7).

$$\mathrm{Exp}(\log(E\mu|\lambda)) = \mathrm{Exp}\left(\alpha + x'\beta\right) = E(\mu|\lambda) = e^{\alpha} + e^{x'\beta} \qquad (32.7)$$

The exponential values for the parameters β_{k} are referred to as multipliers or incidence ratios (IR), since they multiply the expected count or incidence of the outcome by a function of the predictor variable (x). The exponential values of the parameters times the predictors $(e^{x'\beta})$ are typically what is of interest to criminologists since one wants to know how much a predictor variable increases the expected count or incidence of crime. For example, if the expected count is multiplied by 1.0, this means that the variable has no effect. If the expected count is multiplied by 1.5, this means that the expected count increases by a factor of 1.5. If the multiplier is 0.5, then one reduces the expected count by 0.5. Analysts often use relative percentage terms to interpret the multipliers, such that 1.7 would be interpreted as increasing the expected count by 70%, and 0.6 would be interpreted as reducing the expected count by 40%. If you are used to thinking of things in terms of relative ratios, then the IR expression may be more intuitive than percentage changes.

Poisson regression can also be used to describe expected rates, when the rate is a count of crimes divided by a specific unit of exposure (e.g., population, time, etc.). In the context of crime counts, the rate could be an expression of crimes per unit of time. A criminologist, for example, may be interested in modeling the expected number of crimes committed per week or month or number per neighborhood or city population. In Poisson regression, estimating the rate is handled by adding an "offset term" ($\log^*$ exposure) to the right-hand-side of the equation, with the parameter estimate constrained to equal 1 – so that it doesn't get absorbed into the intercept and predictors (Maddala 1983). Offset simply means that the measure is treated differently from the "set" of explanatory variables on the right hand side of the equation. By constraining the estimated offset term to equal 1 gives the equivalent of a rate (Maddala 1983). Equation (32.8) shows that the addition of this offset term results in a model modification to (32.7).

$$\log\left(E(\mu|\lambda)\right) = \log(\text{exposure}) + \alpha + x'\beta \tag{32.8}$$

or equivalently

$$\frac{1}{\log(\text{exposure})} * \log(E(\mu\,|\lambda) = \log(\text{exposure}) + \alpha + x'\beta^* \frac{1}{\log\,(\text{exposure})}$$

$$\log\left(\frac{E(\mu\,|\lambda)}{\text{exposure}}\right) = \alpha + x'\beta$$

We take the logarithm of the exposure variable and constrain the parameter to equal 1 so that expected counts from observations are treated as a fixed scale. For example, suppose one wants to know the expected count of street crimes among a sample of released offenders. In this context, someone who committed four crimes and was out of prison for 4 days would have a rate of 1.0 crime per day. Someone who committed four crimes but was out of prison for 40 days would have a rate of 0.1 crimes per day. The ability to express counts as rates provides an important extension to those interested in crime counts by different units like offenders, populations, or time. Statistical software makes this all easy to do now. STATA, for example, will calculate the expected rate as an exposure variable directly in the estimation so that one doesn't have to convert the variable into is logarithmic form (see Stata version 10.0, 2005). Most statistical packages will allow one to declare an offset variable in estimations. In this case, one needs to log the exposure variable first and then declare it as the offset variable.

COUNT MODELS OF ARRESTS AMONG CALIFORNIA YOUTH AUTHORITY PAROLEES

In the following examples, we estimate Poisson and the related negative binomial and zero inflated Poisson GLM models for counts and rates of arrests for a two cohort sample of California Youth Authority (CYA) parolees for their first 3 years following their release from institutions (see Lattimore et al. 1997, 2004 on samples and measures). We first estimate a standard Poisson regression on total crime counts. We then examine how the linear predictors change when one adjusts for offending rates per street time (number of crimes/number of days an offender was on the street).

For these examples, we predict crime counts using a set of variables that attempt to capture individual offender propensity. Specifically, we measure each individual's problem history according to their number of arrests in the previous 3 years, their age at first officially recorded arrest, whether there was any indication in their probation file of alcohol, drug abuse, gang membership while incarcerated in CYA, previous gang membership prior to incarceration, and recorded acts of violence with incarcerated in CYA. We also include measures of their age upon release, their race or ethnicity, and whether they were from Los Angeles County (given that this is the largest county of referral). The basic descriptive statistics for each measure are shown in Table 32.1.

The distribution of arrests in the first 3 years after release from CYA is displayed in Fig. 32.2. This distribution of arrests is highly skewed to the right, ranging from 0 to 20 but

TABLE 32.1. Descriptive statistics of sample of CYA parolees

Description (variable name)	*n*	Mean	Std. Dev.	Min	Max
Post-release					
Total Arrests 3 years (Arrests3pos)	3,612	2.7267	2.5115	0	20
Days on the street (Daysstreet)	3,612	916.0285	253.0843	0	1095
Cohort (1 = 86/87) (Cohort)	3,612	0.4640	0.4989	0	1
Problem history					
Number of arrests in prior 3 years (Arrests3pre)	3,612	8.0584	5.1910	0	102
Age at first arrest (Agefirst)	3,612	13.5385	2.7095	3	21
Alcohol abuse (0 no, 1 yes) (Alcohol)	3,612	0.6373	0.4808	0	1
Drug abuse (0 no, 1 yes) (Drugabuse)	3,612	0.7674	0.4225	0	1
CYA gang member (0 no, 1 yes) (CYAgang)	3,612	0.2193	0.4138	0	1
Prior gang member (0 no, 1 yes) (Pregang)	3,612	0.3768	0.4846	0	1
CYA violence (0 no, 1 yes) (CYAviolence)	3,612	0.4261	0.4946	0	1
Age at release (years) (Agerel)	3,612	19.3729	2.0160	12	25
White (reference category)	3,612	0.3319	0.4710	0	1
Black (=1)	3,612	0.3735	0.4838	0	1
Hispanic (=1)	3,612	0.2946	0.4559	0	1
Los Angeles County (0 no, 1 yes) (LAcounty)	3,612	0.4236	0.4942	0	1

NOTE: Variables with a range of 0 to 1 are indicator variables

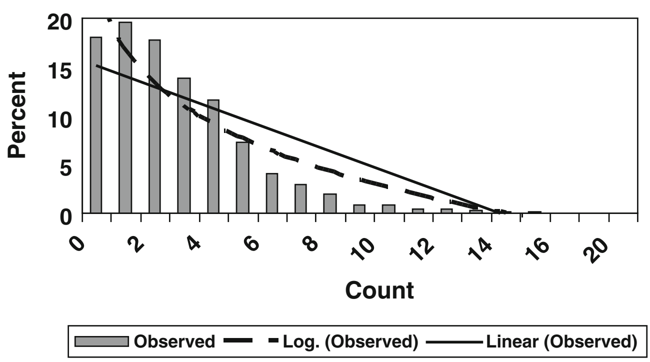

FIGURE 32.2. Distribution of observed crime counts.

TABLE 32.2. Poisson regression of postrelease arrest counts (n = 3,612)

	Coefficient	*Std. Err*	*Z*	*P > |z|*	*95% lower CI*	*95% upper CI*	Exp (B)
Cohort	−0.109	0.022	−4.81	0.000	−0.154	−0.064	0.896
Arrests3pre	0.016	0.001	11.43	0.000	0.013	0.0197	1.01
Agefirst	−0.042	0.004	−10.38	0.000	−0.051	−0.034	0.957
Alcohol	−0.085	0.023	−3.62	0.000	−0.131	−0.039	0.918
Drugabuse	0.182	0.027	6.59	0.000	0.128	0.237	1.20
CYAgang	0.127	0.028	4.49	0.000	0.071	0.183	1.13
Pregang	0.023	0.026	0.86	0.390	−0.029	0.075	1.02
CYAviolence	0.128	0.022	5.85	0.000	0.085	0.172	1.13
Agerel	0.030	0.005	5.40	0.000	0.019	0.041	1.03
Black	0.301	0.027	10.89	0.000	0.247	0.356	1.35
Hispanic	0.209	0.029	7.14	0.000	0.152	0.267	1.23
LAcounty	−0.110	0.023	−4.76	0.000	−0.155	−0.0649	0.895
Constant	0.565	0.105	5.35	0.000	0.358	0.772	

Note: Chi-square model fit = 651.56 (df = 12), $p < 0.001$. Pseudo-$R^2 = 0.04$

with a mean value of 2.72, suggesting that a Poisson distribution may provide an appropriate fit to the data. This distribution, as well as observing the log of the trend line, suggests that this may be an appropriate count to estimate using a Poisson distribution. One can also appreciate the log of the trend line for the distribution is closer to the data than the linear trend – again suggesting the Poisson distribution as a useful approximation.

Table 32.2 shows the results from the Poisson regression model of total arrests during the three years after release. One can see that there are a number of statistically significant predictors. This shouldn't be surprising given the relatively large sample size of over 3,000 observations. In fact, if we examine the expected counts from the exponential value of the coefficients (expB), we see that most of the predictors produce relatively small changes in the expected count of arrests. There are some notable exceptions. For example, having a history of drug abuse (Drugabuse) multiplies the expected count by 1.20 or increases it by 20%. Similarly, membership in a gang (CYAgang) and violent behaviors (CYAvio) while detained in CYA facilities, both multiply the expected counts by 1.13 or 13%. There are also differences between Blacks and Hispanics and Whites, but we don't interpret these coefficients because we suspect that these variables may be serving as proxies for variables that are not included in the equation such as household income.

If we are interested in examining how these predictors are related to offending rates and not counts, we can accomplish this by adjusting our estimates for Daysstreet, the time individuals in our sample spent on the street and were, thus, "at risk" of committing crimes and being arrested. For example, parolee A with 10 offenses but who is on the street for 1 month would be a higher rate offender than parolee B who commits 10 offenses while on the street for an entire year. In this example, parolee A would have an offense rate of 0.33 crimes per day (10 crimes/30.4 days), whereas parolee B would have an offense rate of 0.03 crime per day (10/365 days). Parolee B's offense rate is only 1/12th that of parolee A!

Table 32.3 compares the multipliers or incidence ratios (Exp(B)) that were presented in Table 32.2 with the multipliers generated when the same set of predictors are included and we adjust for the days that each parolee was free on the street. This is accomplished by entering the log of the number of days free on the street into the Poisson regression equation as

TABLE 32.3. Comparisons of incidence ratios with and without adjustments for street time

Variable	Rates (street time adjustment) Exp(B)	Counts (no street time adjustment) Exp(B)
Cohort	0.875***	0.896***
Arrests3pre	1.018***	1.017***
Agefirst	0.951***	0.958***
Alcohol	0.922***	0.918***
Drugabuse	1.236***	1.201***
CYAgang	1.179***	1.136***
Pregang	1.054*	1.023
CYAviolence	1.141***	1.138***
Agerel	1.010*	1.031***
Black	1.408***	1.353***
Hispanic	1.225***	1.233***
LAcounty	0.884***	0.895***
Observations	3,609	3,612

Note: Chi-square model fit = 880.17 (df = 12), $p < 0.001$. Pseudo-$R^2 = 0.05$
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

an offset variable whose coefficient is fixed to equal one or, where allowed by the statistical package such as STATA declaring an exposure variable that is entered directly into the log-link function. In both estimations, it is important to remember that street time is NOT a predictor variable and should not be absorbed into the direct equation estimation, rather it is an adjustment term (see (32.8) to convince yourself of this). As can be seen in Table 32.3, the results from the model adjusting for street time are materially the same as those presented in Table 32.2, where we did not adjust for street time. In fact, the coefficients change very little. The calculation of rate, however, is useful because it puts every observation's count of arrests on the same scale. In terms of interpretation, we can interpret the multipliers here as increasing the expected count of crime per day. A past history of drug abuse (Drugabuse), for example, multiplies the expected count of crimes per day by 1.236 or increases it by 23.6%, whereas it increases the expected total count over the 3 years post release by 20.1% (IR = 1.201).

Now that we have seen how to interpret the coefficients of predictors of counts and rates of counts from a Poisson model, we need to address the next question of how well the Poisson model approximates the data. Remember that the expected mean and variance from the Poisson distribution should be equal. If we compare the predicted mean and its variance from these models, we find that the expected mean is 2.72, which is the same as the observed mean; but, the estimated variance is less than the mean for both models (0.55 for the count model and 1.24 for the rate model that adjusts for street time).[1] This finding suggests that the conditional variance is underdispersed in our Poisson regression estimations.

The results suggesting that the observed data are underdispersed compared to the dispersion one would expect with a Poisson regression model also suggest that we may want to

[1] Most statistical packages have a default routine to save predicted values. Remember this is simply the value of the expected count derived from the model. In STATA there is a default command after model estimations that will allow you to obtain the predicted counts.

estimate a zero-inflated Poisson model in an effort to account for the fact that 18% ($n = 654$) of the parolees in our sample are not observed committing offenses. This group of observations may be driving the variance of the conditional distribution downward. An alternative possibility is that we don't have the right set of predictors to estimate the Poisson model, and if we did have the right set of predictors, the conditional variance would inflate to equal the conditional mean. In any case, we will assume that we have the correct predictors and that the conditional distribution of choice should account for the zero crime counts in our data.

Table 32.4 presents the results from the zero inflated Poisson (ZIP) model and the negative binomial (NB) regressions of crime counts. Column 1 presents the coefficients from the ZIP of the expected counts and column 2 list the inflation estimates which are estimates of the expected zeros of the distribution. Notice that the inflation estimates are in the opposite direction from the expected counts – which is what you would expect, i.e., predictors of positive arrest counts are negatively predictive of zero arrests. As one can see, a past history of drug abuse (Drugabuse) multiples the expected count by 1.13 (increases counts by 13%), but multiplies the expected probability of zero arrests by 0.68 or reduces the expected probability of zero arrests by 32%. There is some noticeable change in the estimated effect of a few predictors when one accounts for the zeros in the empirical distribution. For example, the effect on the expected count of arrests of gang membership in the CYA results in a multiplier of 0.059 in the ZIP model but a multiplier of 1.136 in the standard Poisson model. Similarly, Drugabuse multiplies the expected count of arrests by 1.133 in the ZIP model but 1.201 in the standard Poisson regression model. While these aren't substantial differences, they do provide one with evidence that on the margin the various models could lead to different conclusions.

TABLE 32.4. Zero inflated Poisson (ZIP) and negative binomial (NB) regression of arrests

Variables	ZIP (1) Arrests3pos Exp(B)	(2) Inflate Exp(B)	NB (3) Arrests3pos Exp (B)
Cohort	0.924***	1.409***	0.872***
Arrests3pre	1.012***	0.911***	1.026***
Agefirst	0.973***	1.123***	0.956***
Alcohol	0.926***	1.051	0.922**
Drugabuse	1.133***	0.680**	1.203***
CYAgang	1.059*	0.536***	1.126***
Pregang	1.025	1.064	1.021
CYAviolence	1.112***	0.818	1.144***
Agerel	1.050***	1.121***	1.036***
Black	1.247***	0.476***	1.379***
Hispanic	1.243***	1.091	1.229***
LAcounty	0.897***	0.989	0.906***
Constant	1.310**	0.0104***	1.523**
Alpha			0.400***
Observations	3,612	3,612	3,612

Note: Alpha = dispersion parameter likelihood ratio test that Alpha = 0 was 1179.07 ($p < 0.001$).
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

It is often the case in criminology that the counts are so skewed to the right that the conditional variance will exceed the conditional mean (see Osgood 2000 for examples with youth robbery arrests). Under such circumstances, the conventional approach is to estimate a negative binomial regression model. We follow such an approach here for an illustration, not because the data are overdispersed. The results from the NB regression of the counts of arrests are shown in Table 32.4, column 3. The results indicate that the coefficients are similar to the ZIP specification. Column 4 also shows the dispersion parameter, alpha, and indicates that is significantly smaller than 1, indicating that the data are indeed underdispersed. If the dispersion parameter was significantly >1, one would argue that there is evidence of overdispersion.

MODEL DIFFERENCES

Now that we have seen how the Poisson, ZIP, and NB regression models differ, the question arises of which model to use? Unfortunately, there is no correct answer. Specifically, the best model depends on the data and the set of predictors available. One method for choosing between the alternatives is to compare dispersions for each conditional model and examine their relative fit to the data. How dispersed are the data from the conditional mean in each model and how well does each model fit the data? The main problem with this approach is that it assumes that the predictors are correct and complete, providing a basis for comparison and the selection of the right model. In criminology, this is rarely the case; however, we will demonstrate such an approach as a method of model comparison.

Table 32.5 displays the conditional means and variances for the Poisson, ZIP, and NB models, with and without adjustments for street time. Adjusting for street time will inflate the variance by equivalent metrics, such that the dispersion is still less than the mean for the Poisson. One can see that adjusting for street time increases the expected variance for the Poisson model so that it is close to the unadjusted NB model. With the exception of the NB model that adjusts for street time, all models have conditional variances that are substantially lower than their conditional means. This would appear to present prima facie evidence in favor of the NB models. Again, if one assumes that we have all of the correct predictor variables, which is not a reasonable assumption. In fact, given that the coefficients do not change materially it is probably wiser to use the Poisson model than any of the other variants for the simple reason that it provides the most parsimonious approximation and make no assumptions about the error distribution (see Berk and MacDonald 2008 for full exposition).

An alternative approach is described in full in Long (1997) and Long and Freese (2001). This approach involves comparing the fit of each model against their respective probability distribution. In this case, we compare the fit of the estimated Poisson regression model against the theoretical Poisson distribution, the estimated ZIP regression model against the theoretical ZIP distribution, and the NB regression model against the NB distribution. Figure 32.3 displays the fit of each model compared to its respective distribution. These graphs suggest

TABLE 32.5. Conditional means and variances from estimated models

	Poisson	*P_Street*	*NB*	*NB_Street*	*ZIP*	ZIP_Street
Mean	2.72	2.72	2.74	3.01	2.72	2.80
Variance	0.558	1.23	1.25	3.32	0.534	1.20

that the predicted values from the Poisson regression model underpredicts the lower counts compared to its theoretical Poisson distribution. Given that there were 18% of cases with zero counts of crime, this shouldn't be a big surprise.

Long and Freese (2001) also provide some useful graph commands that allows us to plot the average count from each predicted model against a theoretical distribution. For example, if we select the Poisson distribution as our distribution of choice, we see in Fig. 32.4 how much the expected average count from each of our models deviate from the theoretical Poisson distribution. One can see that the Poisson regression model's deviation (devpois) appears to be underperforming relative to the ZIP (devzip) and NB (devnbreg) models with respect to the lower counts.

DIAGNOSTICS AND RELATED ISSUES

Both the NB and ZIP variants of the Poisson regression model place stringent assumptions with respect to the process generating the excess or dearth of variation relative to the Poisson. Berk and MacDonald (2008) provide clear examples of excess variation and the assumptions imposed by the negative binomial model. In simple terms, both the NB and ZIP assume there is no omitted variable bias driving the over- or under-dispersion. Rarely is it the case in observational studies that one has measured all the relevant predictors. It is useful, therefore, to compare outcomes under both the Poisson and other variants to see how sensitive the findings are to these corrections. In practice, it may be advisable to use the Poisson and work on getting the systematic structure of the model correct before proceeding to other corrections. If the average counts of outcomes are high enough (e.g., 20 or greater) it may be advisable to just use OLS.

Regression diagnostics rarely appear in criminology publications that use Poisson models or the other variants. But the basic assumptions of GLM models regarding collinear predictors and influential observations or outliers on parameter estimates hold for count models as well. Since Poisson regression and its variants are GLM models, the same basic assumptions have to be made that one relies on with other classes like OLS. For example, the negative binomial distribution has an extreme right skew which allows the negative binomial regression model to incorporate particularly high counts, but high counts can also be important outliers in negative binomial models with an observational dataset. In other words, outliers can drive the parameter estimates in count models, as well as traditional OLS. The usual diagnostics for linear regression models can be used to see how well count models hold up. Unfortunately, most statistical software packages don't provide standard regression diagnostics like Cooks D, Dfbetas, and residual plots as default post estimation routines. Residuals can, however, be easily calculated by subtracting the observed counts from those predicted by the model, i.e. $E(Y\text{i}) - Y\text{i}$. One can then inspect residuals for outliers and influential observations. For example, what happens to the model when one drops observations greater than 2.0 or 3.0 standard deviations from the expected mean? One also can easily calculate variance inflation factors (VIF) by regressing each predictor variable on each other and examining the square of the correlation coefficient (R-square) by the following equation: VIF $= 1/1$-R-square. VIF values >4.0 may indicate a problem with multicollinearity among the predictors.

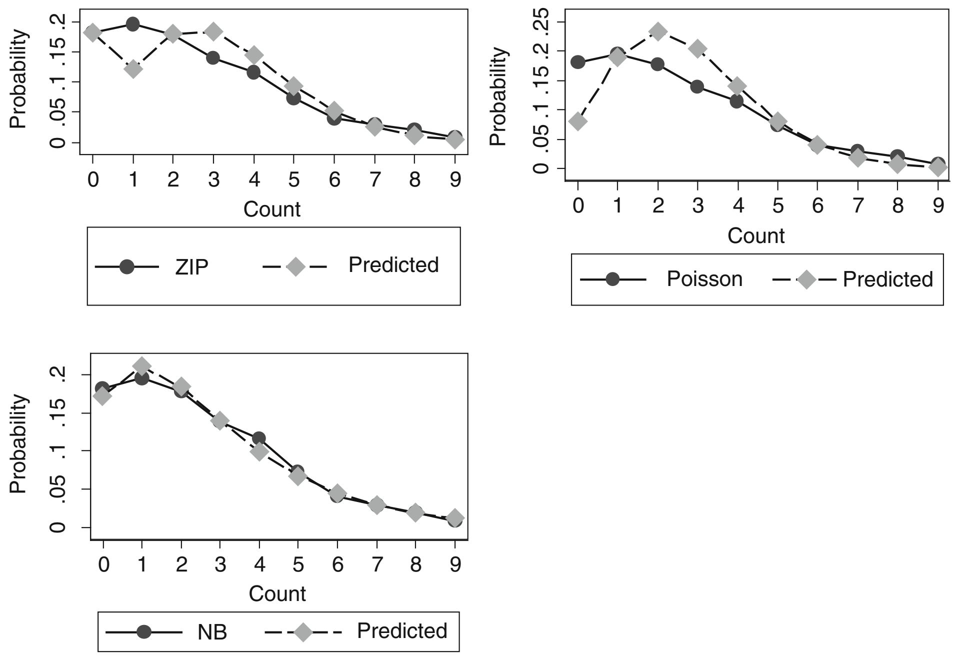

FIGURE 32.3. Predicted vs. theoretical distributions.

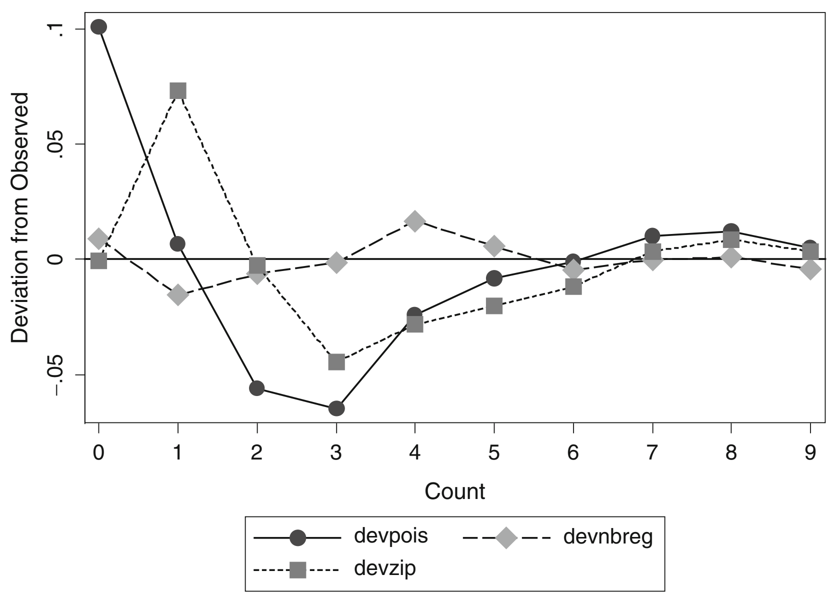

FIGURE 32.4. Predicted values versus Poisson distribution.

CONCLUSIONS

Count models are important to criminology, as so many of the outcomes of interest, such as arrests, take the form of counts in which small counts (and averages) often dominate. Now, most statistical computer software packages offer estimation routines for Poisson, NB and ZIP models, greatly facilitating their use. As is true with all statistical models, the analyst should consider the assumptions underlying the selected model and assess whether these assumptions are reasonable given the data and theory under consideration. Users should follow basic guidelines of regression diagnostics and apply these procedures when they have count outcomes, just as they would with continuous data. The popularity of Poisson regression models has resulted in their use with less focus on their functional utility. For readers with more interest in the limitations of the Poisson regression model and the negative binomial variant, see Berk and MacDonald (2008).

There are several extensions of the class of GLM models for the Poisson that have been developed and applied to good end in criminology. Among these extensions are mixture models that allow one to specify or partition the variance in the Poisson process at different levels of analysis, such as between individuals and the neighborhoods, schools, or prisons under which they are located, or the trajectories or clusters of offenders according to the homogeneity of semi-homogenous groups (Nagin and Land 1993; Nagin 2005; Gellman and Hill 2008; Kreuter and Muthen 2008). The extensions to the GLM of Poisson are discussed in detail elsewhere in this text and offer more flexibility in specification, because they allow the user to test explicit theories that argue the propensity for crime is conditional on different units of analysis. To make appropriate inference with these extensions of the Poisson regression model, it is important that one have a reasonable theory or experiment by which one can attribute independent assignment across different units. For example, within a sample of active offenders, there may exist a subgroup that lived in a set of neighborhoods subjected to a particular police intervention. Under such an example, one could distinguish the covariates for the individual propensity for offending from that generated by the neighborhood-level police intervention. Here, one could use a multilevel Poisson model (commonly referred to among criminologists as a hierarchical linear model). Similarly, there may exist a subgroup of offenders that prefer specific offenses (e.g., burglary vs. violence) and one could estimate Poisson regressions conditional on different offending trajectories. If one has a strong theoretical reason to believe that the process generating the data differs by different units (neighborhood, crime type, or type of offender) one can extend the Poisson regression model to a variety of mixture models. There is also the potential for imbedding a counterfactual model of causal inference using the expected counts as propensities and matching on distance between expected counts or rates between groups of interest. Under such an extension, one could compare the expected count of crimes after each group has been exposed to some treatment. Haviland et al. (2008) provide an example of this by matching on expected crime count propensities for youth prior to joining gangs to compare the difference in subsequent criminal acts between those who join gangs with a group with similar criminal propensities.

Finally, in the absence of compelling evidence, simpler models may be better. As noted earlier, if the average count in the data is large (e.g., >20), OLS may provide an adequate and, likely, best fit to the data. Similarly, Poisson regression may also provide an adequate and, perhaps, best fit even when there is apparent evidence of under- or over-dispersion if the user is not convinced that all (important) parameters have been included in the model. As demonstrated here, there are a number of approaches for examining the fit of alternative models and a thorough analysis should incorporate these.

REFERENCES

Agresti A (2007) An introduction to categorical data analysis, 2nd edn. Wiley, New York

Angrist J (2006) Instrumental variables methods in experimental criminological research: what, why and how. J Exp Criminol 2:23–44

Berk R, MacDonald J (2008) Overdispersion and poisson regression. J Quant Criminol 24:269–284

Blumstein A, Cohen J, Roth JA, Visher CA (eds) (1986) Criminal careers and "career v criminals," vol 2. National Academy Press, Washington DC

Braga A (2003) Serious youth gun offenders and the epidemic of youth violence in Boston. J Quant Criminol 19: 33–54

Cameron AC, Trivedi PK (1998) Regression analysis of count data. Cambridge University Press, Cambridge

Gellman A, Hill J (2008) Data analysis using regression and multilevel/hierarchical models, 5th edn. Cambridge University Press, Cambridge

Good IJ (1986) Some statistical applications of poisson's work. Stat Sci 1:157–170

Greene WH (1997) Econometric analysis, 3rd edn. Prentice Hall, Upper Saddle River, NJ

Greenberg DF (1991) Modeling Criminal Careers. Criminology 29:17–46

Haviland AM, Nagin D, Rosenbaum P, Tremblay R (2008) Combining group-based trajectory modeling and propensity score matching for causal inferences in nonexperimental longitudinal data. Developmental Psychology 44(2):422–436

Kreuter F, Muthen B (2008) Analyzing criminal trajectory profiles: bridging multilevel and group-based approaches using growth mixture modeling. J Quant Criminol 24:1–31

Lattimore PK, Krebs C, Koetse W, Lindquist C, Cowell A (2005) Predicting the effect of substance abuse treatment on probationer recidivism J Exp Criminol 1:159–189

Lattimore PK, Linster RL, MacDonald JM (1997) Risk of death among serious young offenders. J Res Crime Delinq 34:187–209

Lattimore PK, MacDonald JM, Piquero AR, Linster RL, Visher CA (2004) Studying frequency of arrest among paroled youthful offenders. J Res Crime Delinq 41:37–57

Long JS (1997) Regression models for categorical and limited dependent variables. Sage Publications, Thousand Oaks, CA

Long JS, Freese J (2001) Regression models for categorical dependent variables using stata. Stata Press, College Station, TX

Maddala GS (1983) Limited-dependent and qualitative variables in econometrics. Cambridge University Press, New York

Gardner W, Mulvey EP, Shaw ES (1995) Regression analyses of counts and rates: poisson, overdispersed poisson, and negative binomial models. Psychol Bull 118:392–404

Nagin DS (2005) Group-based modeling of development. Harvard University Press, Cambridge, MA

Nagin DS, Land KC (1993) Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed poisson model. Criminology 31:327–362

Osgood W (2000) Poisson-based regression analysis of aggregate crime rates. J Quant Criminol 16:21–43

Osgood DW, Chambers JM (2000) Social disorganization outside the metropolis: an analysis of rural youth violence. Criminology 38:81–115

Piquero AR, Farrington DP, Blumstein A (2003) The criminal career paradigm: background and recent developments. In: Tonry M (ed) Crime and justice: a review of research, vol 30. University of Chicago Press, Chicago, pp 359–506

Parker K (2004) Industrial shift, polarized labor markets and urban violence: modeling the dynamics between the economic transformation and disaggregated homicide. Criminology 42:619–645

Paternoster R, Brame R, Bachman R, Sherman L (1997) Do fair procedures matter? The effect of procedural justice on spouse assault. Law Soc Rev 31:163–204

Sampson RJ, Laub JH (1996) Socioeconomic achievement in the life course of disadvantaged men: military service as a turning point, circa 1940–1965. Am Sociol Rev 61:347–367

Stigler S (1986) The history of statistics: the measurement of uncertainty before 1900. The Belknap Press of the Harvard University Press, Cambridge, Massachusetts