Journal of Quantitative Criminology, Vol. 16, No. 1, 2000

Poisson-Based Regression Analysis of Aggregate Crime Rates

D. Wayne Osgood [1]

bution as a tool for resolving common problems in analyzing aggregate crime rates. When the population size of an aggregate unit is small relative to the offense rate, crime rates must be computed from a small number of offenses. Such data are ill-suited to least-squares analysis. Poisson-based regression models of counts of offenses are preferable because they are built on assumptions about error distributions that are consistent with the nature of event counts. A simple elaboration transforms the Poisson model of offense counts to a model of per capita offense rates. To demonstrate the use and advantages of this method, this article presents analyses of juvenile arrest rates for robbery in 264 nonmetropolitan counties in four states. The negative binomial variant of Poisson regression effectively resolved difficulties that arise in ordinary least-squares analyses.

KEY WORDS: Poisson; negative binomial; crime rates; aggregate analysis.

1. INTRODUCTION

The purpose of this paper is to introduce a statistical approach to analyzing aggregate crime rates that solves problems arising from small populations and low base-rates. In aggregate analyses, the units of the sample are aggregations of individuals, such as neighborhoods, cities, and schools, and the researcher is interested in explaining variation in crime rates across those units. These crimes rates are defined as the number of crime events (e.g., arrests, victimizations, crimes known to the police) divided by the population size, often reported as crimes per 100,000.

The standard approach to analyzing per capita rates such as these is to use the computed rates for each aggregate unit (or a transformed version of them) as the dependent variable in an ordinary least-squares (OLS) regression. For reasons discussed below, this least-squares approach is inappropriate when rates for many of the units must be computed from small

[1] Crime, Law and Justice Program, Department of Sociology, 201 Oswald Tower, Pennsylvania State University, University Park, Pennsylvania 16802. e-mail: wosgood@psu.edu.

0748-4518/00/0300-0021$18.00/0

numbers of events. The present paper demonstrates how to resolve this problem through Poisson-based regression models that are well suited to such data. The statistical basis of this analytic approach is well established (e.g., Cameron and Trevedi, 1998; Gardner *et al.*, 1995; King, 1989; Liao, 1994; McCullagh and Nelder, 1989; Land *et al.*, 1996). In criminology, Poisson-based regression models have become common in analyses of criminal careers (Greenberg, 1991; Land *et al.*, 1996; Nagin and Land, 1993; Rowe *et al.*, 1990), but they have rarely been applied to aggregate analysis of crime or other social phenomena (for examples see Bailey *et al.*, 1994; Sampson *et al.*, 1997). In the hope of making these techniques accessible to a broader range of researchers, the present article is devoted to articulating the special advantages of these models for solving problems in analyzing aggregate data.

1.1. The Problem

Because arrests are discrete events, the possible crime rates for any given population size are those corresponding to integer counts of crimes. For instance, in a county of 200,000 individuals, every additional crime increases the crime rate by half an arrest per 100,000, while in a neighborhood of 5000 each crime corresponds to 20 crimes per 100,000. If the population sizes of the aggregate units are large relative to the average arrest rate, then the calculated rates will be sufficiently fine-grained that there is no harm in treating them as though they were continuous and applying least squares statistics. For almost any measure of offending, populations of several hundred thousand should prove adequate in this regard. When populations are small relative to offense rates, however, the discrete nature of the crime counts cannot be ignored. Indeed, for a population of a few thousand, even a single arrest for rape or homicide might correspond to a high crime rate. Low counts of crimes are common for offense-specific analyses, samples of small towns and rural areas, and analyses of subpopulations, such as females versus males or specific age categories.

Crime rates based on small counts of crimes present two serious problems for least squares analysis. First, because the precision of the estimated crime rate depends on population size, variation in population sizes across the aggregate units will lead to violating the assumption of homogeneity of error variance. We must expect larger errors of prediction for per capita crime rates based on small populations than for rates based on large populations. Second, normal or even symmetrical error distributions of crime rates cannot be assumed when crime counts are small. The lowest possible crime count is zero, so the error distribution must become increasingly skewed (as well as more decidedly discrete) as crime rates approach this lower bound. As populations decrease, an offense rate of zero will be observed for a larger and larger proportion of cases. Thus, there is an effective censoring at zero that is dependent on sample size, which has considerable potential for biasing the resulting regression coefficients.

The standard solution to the problems of low offense counts has been to increase the level of aggregation, such as analyzing only large cities or combining specific offenses into broad indices. For instance, one rarely see analyses of homicide for populations less than several hundred thousand. Not only does this strategy preclude analyses about many interesting questions, it leads to coarser measurement of important explanatory variables, such as being forced to assume that a single poverty rate applies equally well to all neighborhoods in a city. Fortunately, there is an alternative data analytic approach that resolves these problems.

2. THE POISSON REGRESSION MODEL

The Poisson distribution has been useful for many problems in criminology and criminal justice. Indeed, Poisson originally derived the distribution for analyzing rates of conviction in France during the 1820s (see Maltz, 1994). Maltz (1994) reviews many uses of the Poisson distribution for modeling phenomena related to crime, such as assessing the potential for selective incapacitation, projecting prison populations, and estimating the size of the criminal population. The present paper focuses specifically on Poisson-based regression models, which relate explanatory variables to dependent variables that are counts of events. These models can solve the problems described above because they allow us to recognize the dependence of crime rates on *counts* of crimes. Several good, nontechnical descriptions of Poisson regression are now available (e.g., Gardner *et al.*, 1995; Liao, 1994; Land *et al.*, 1996), so my description of these models is brief, emphasizing the features most relevant to aggregate analysis.

The Poisson distribution characterizes the probability of observing any discrete number of events (i.e., 0, 1, 2, . . .), given an underlying mean count or rate of events, assuming that the timing of the events is random and independent. For instance, the Poisson distribution for a mean count of 4.5 would describe the proportion of times that we should expect to observe any specific count of robberies (0, 1, 2, . . .) in a neighborhood, if the "true" (and unchanging) annual rate for neighborhood were 4.5, if the occurrence of one robbery had no impact on the likelihood of the next, and if we had an unlimited number of years to observe. Figure 1 shows the Poisson distribution for four mean counts of arrests. When the mean arrest count is low, as is likely for a small population, the Poisson distribution is skewed, with only a small range of counts having a meaningful probability of occurrence. As the mean count grows, the Poisson distribution increasingly approximates the normal. The Poisson distribution has a variance equal to the mean count. Therefore, as the mean count increases, the probability of observing any specific number of events declines and a broader range of values have meaningful probabilities of being observed.

Our interest is in per capita crime rates rather than in counts of offenses, and Fig. 2 demonstrates the correspondence between rates and counts. Figure 2 translates the Poisson distributions of crime counts in Fig. 1 to distributions of crime rates. Given a constant underlying mean rate of 500 crimes per 100,000 population, population sizes of 200, 600, 2000, and 10,000 would produce the mean crime counts of 1, 3, 10, and 50 used in Fig. 1. For the population of 200, only a very limited number of crime rates are probable (i.e., increments of 500 per 100,000), but those probable rates comprise an enormous range. As the population base increases, the range of likely *crime rates* decreases, even though the range of likely *crime counts* increases. The standard deviation around the mean rate shrinks from 500 crimes per 100,000 for a population of 200 to 71 crimes per 100,000 for a population of 10,000. Thus, Fig. 2 illustrates the effect of population size on the accuracy of estimated crime rates.

The basic Poisson regression model is

$$\ln(\lambda_i) = \sum_{k=0}^{K} \beta_k x_{ik} \tag{1}$$

$$P(Y_i = y_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \tag{2}$$

Equation (1) is a regression equation relating the natural logarithm of the mean or expected number of events for case i, $\ln(\lambda_i)$, to the sum of the products of each explanatory variable, x_{ik}, multiplied by a regression coefficient, β_k (where β_0 is a constant multiplied by 1 for each case). Equation (2) indicates that the probability of y_i, the observed outcome for this case, follows the Poisson distribution (the right-hand side of the equation) for the mean count from Eq. (1), λ_i. Thus, the expected distribution of crime counts, and corresponding distribution of regression residuals, depends on the fitted mean count, λ_i, as illustrated in Fig. 1. The role of the natural logarithm in Eq. (1) is comparable to the logarithmic transformation of the dependent variable that is common in analysis of aggregate crime rates. In both cases, the regression coefficients reflect proportional differences in rates. Liao (1994) provides a detailed discussion of the interpretation of regression coefficients from Poisson-based models.

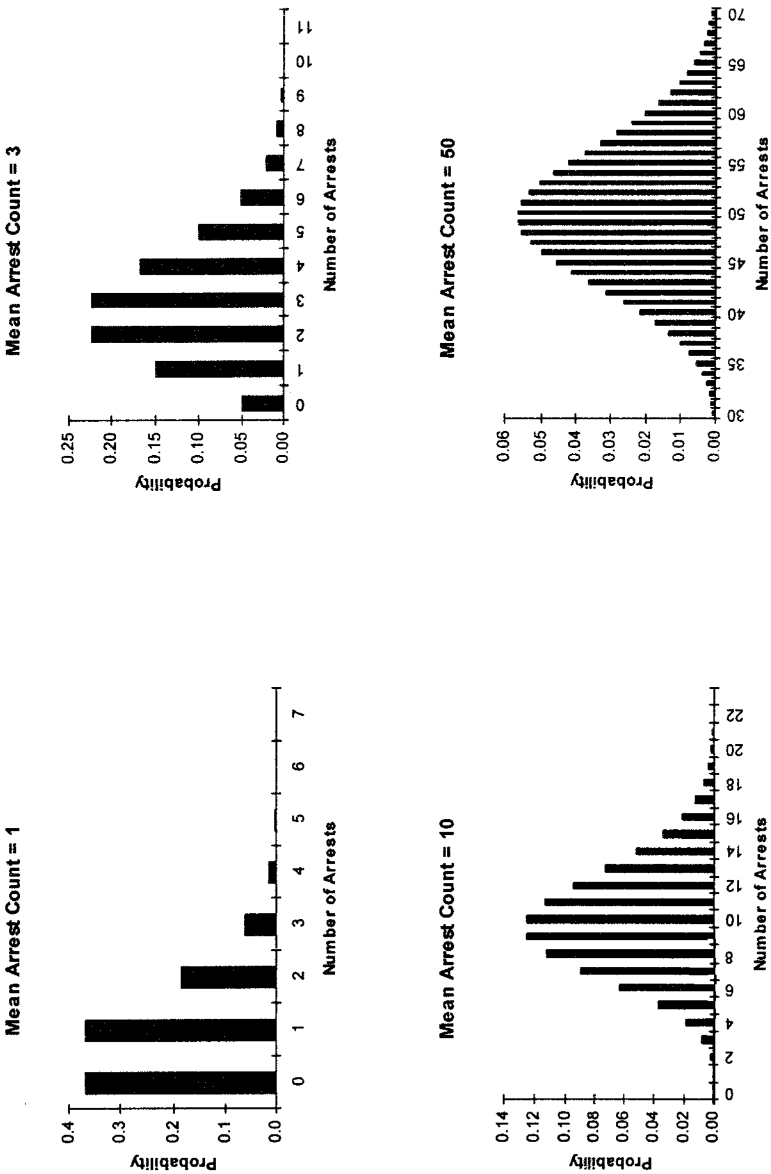

Fig. 1. Poisson distribution for four mean arrest counts.

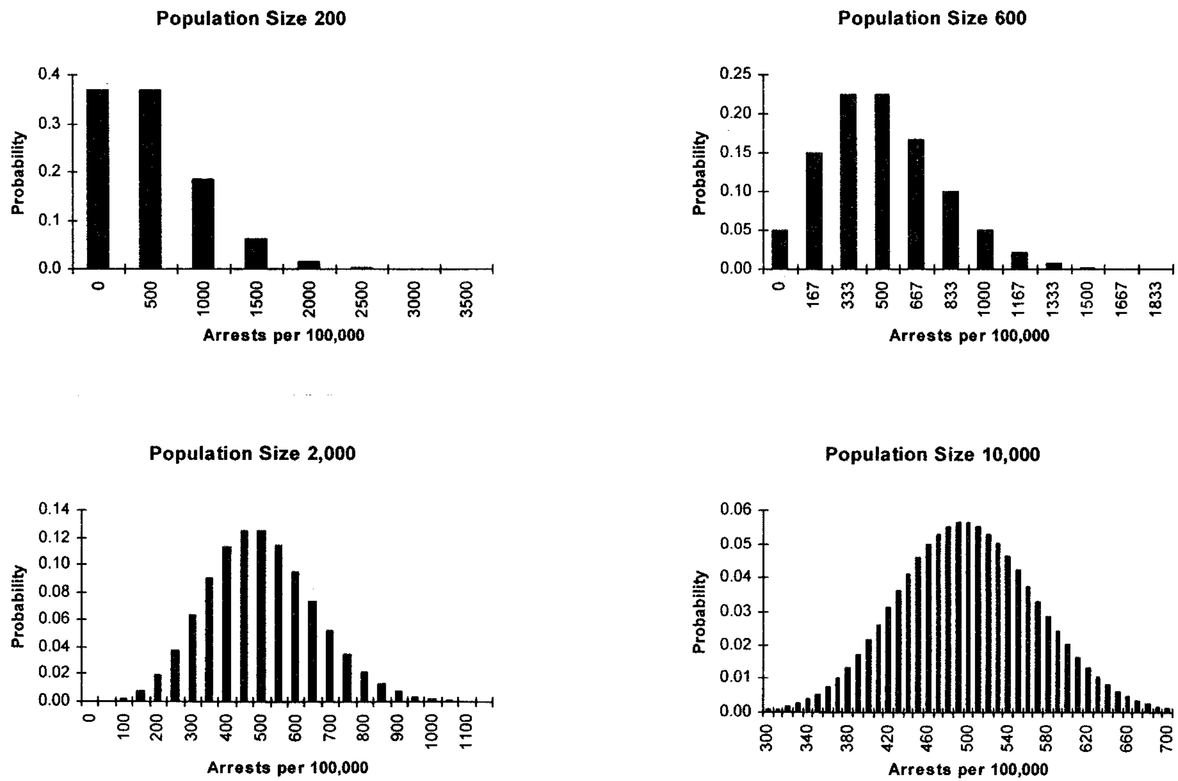

Fig. 2. Poisson distribution of arrest rates for four population sizes, given a mean rate of 500 arrests per 100,000 population.

Next we must alter the basic Poisson regression model so that it provides an analysis of per capita crime rates rather than counts of crimes. If λ_i is the expected number of crimes in a given aggregate unit, then λ_i/n_i would be the corresponding per capita crime rate, where n_i is the population size for that unit. With a bit of algebra, we can derive a variation of Eq. (1) that is a model of per capita crime rates:

$$\ln\left(\frac{\lambda_i}{n_i}\right) = \sum_{k=1}^{k} \beta_k x_{ik}$$

$$\ln(\lambda_i) = \ln(n_i) + \sum_{k=0}^{K} \beta_k x_{ik} \tag{3}$$

Thus, by adding the natural logarithm of the size of the population at risk to the regression model of Eq. (1), and by giving that variable a fixed coefficient of one, Poisson regression becomes an analysis of rates of events per capita, rather than an analysis of counts of events. The same strategy can be used to standardize event count models for other sources of variation across cases, such as the length of the period of observation. Accordingly, computer programs for Poisson regression routinely incorporate this feature.

A Poisson-based regression model that is standardized for the size of the population at risk acknowledges the greater precision of rates based on larger populations, thus addressing the problem of heterogeneity of error variance discussed above. This becomes apparent when we translate the known variance of the Poisson distribution to a standard deviation for the corresponding crime rates. Because the variance of the Poisson distribution is the mean count, λ, its standard deviation will be $SD_\lambda = \sqrt{\lambda}$. The mean count of crimes, in turn, equals the underlying per capita crime rate, C, times the size of the population: $\lambda = Cn$. When a variable is divided by a constant, its standard deviation is also divided by that constant. Therefore, it follows that the standard deviation of a crime rate, computed from a population of size n, will be

$$SD_C = \frac{\sqrt{\lambda}}{n} = \frac{\sqrt{Cn}}{n} = \frac{\sqrt{C}\sqrt{n}}{n} = \frac{\sqrt{C}}{\sqrt{n}}$$

This equation shows that, in the expected distribution of observed crime rates around the fitted mean crime rates produced by Eq. (3), the standard deviation is inversely proportional to the square root of the population size. Thus, Poisson regression analysis explicitly addresses the heterogeneous residual variance that presented a problem for OLS regression analysis of crime rates.

2.1. Overdispersion and Variations on the Basic Poisson Regression Model

The basic Poisson regression model is appropriate only if the probability model of Eq. (2) matches the data. Equation (2) requires that the residual variance be equal to the fitted values, λ_i, which is plausible only if the assumptions underlying the Poisson distribution are fully met by the data. One assumption is that λ_i is the true rate for each case, which implies that the explanatory variables account for all of the meaningful variation among the aggregate units. If not, the differences between the fitted and true rates will inflate the variance of the residuals. It is very unlikely that this assumption will be valid, for there is no more reason to expect that a Poisson regression will explain all of the variation in the true crime rates than to expect that an OLS regression would explain all variance other than error of measurement.

Residual variance will also be greater than λ_i if the assumption of independence among individual crime events is inaccurate. Dependence will arise if the occurrence of one offense generates a short-term increase in the probability of another occurring. For aggregate crime data, there are many potential sources of dependence, such as an individual offending at a high rate over a brief period until being incarcerated, multiple offenders being arrested for the same incident, and offenders being influenced by one another's behavior. These types of dependence would increase the year-to-year variability in crime rates for a community beyond λ_i, even if the underlying crime rate were constant.

For these two reasons, "overdispersion" in which residual variance exceeds λ_i is ubiquitous in analyses of crime data. Applying the basic Poisson regression model to such data can produce a substantial underestimation of standard errors of the β's, which in turn leads to highly misleading significance tests. There are several ways to allow for the possibility of overdispersion (Cameron and Trevedi, 1998; Land *et al*., 1996). Perhaps the simplest is the quasi-likelihood approach (Gardner *et al*., 1995), which retains coefficient estimates from the basic Poisson model but adjusts standard errors and significance tests based on the amount of overdispersion. Other approaches explicitly incorporate a source of overdispersion in the probability model, typically by adding a case-specific residual term to the regression model [Eq. (1) or (3)], comparable to the error term in OLS regression. These versions of Poisson regression are distinguished by the specific assumptions made about the distribution of the residual variation in underlying rates, which may be continuous or discrete (Cameron and Trevedi, 1998).

We illustrate this approach with the negative binomial regression model, which is the best known and most widely available Poisson-based

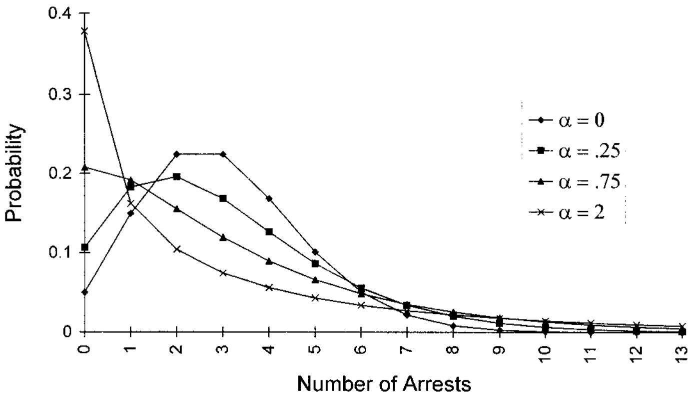

Fig. 3. Negative binomial distributions with mean count of 3, for four levels of residual variance.

regression model that allows for overdispersion. Negative binomial regression combines the Poisson distribution of event counts with a gamma distribution of the unexplained variation in the underlying or true mean event counts, λ_i. This combination produces the negative binomial distribution, which replaces the Poisson distribution of Eq. (2). The formula for the negative binomial is

$$P(Y_i = y_i) = \frac{\Gamma(y_i + \phi)}{y_i!\Gamma(\phi)} \frac{\phi^{\phi}\lambda_i^{y_i}}{(\phi + \lambda_i)^{\phi - y_i}} \tag{4}$$

where Γ is the gamma function (a continuous version of the factorial function), and ϕ is the reciprocal of the residual variance of underlying mean counts, α (Gardner *et al.*, 1995, p. 400).

Figure 3 demonstrates the impact of residual variance on the resulting distribution for a mean count of three crimes. With α equal to zero, we have the original Poisson distribution. For the Poisson, 5.0% of cases would have zero crimes and 1.2% would have eight or more crimes. As α increases, the distribution becomes more decidedly skewed as well as more broadly dispersed. Even for a moderate α of 0.75, the change from the Poisson is dramatic: 20.8% of cases would have zero crimes and 8.8% would have eight or more crimes.

In negative binomial regression (as in almost all Poisson-based regression models), the substantive portion of the regression model remains Eq. (1) for crime counts or Eq. (3) for per capita crime rates. Thus, though the response probabilities associated with the fitted values differ from the basic Poisson regression model, the interpretation of the regression coefficients does not.

3. AN EXAMPLE: SOCIAL DISORGANIZATION AND RURAL YOUTH VIOLENCE

I illustrate the use of Poisson-based regression to study aggregate crime rates with an analysis of rates of juvenile violence in nonmetropolitan counties of four states. This is an elaboration of part of the results presented by (Osgood and Chambers, 2000), and the present article is intended as a methodological companion to that article. Osgood and Chambers (2000) provide a rationale for these analyses in terms of social disorganization theory, offer a full description of the sample and measures, and present analyses for a variety of specific offenses.

The sample consists of the 264 nonmetropolitan counties of Florida, Georgia, South Carolina, and Nebraska, which have total populations ranging from 560 to 98,000. The average population of these counties is roughly 10,000, which is comparable to average neighborhood populations in research comparing neighborhoods within urban centers (Sampson *et al.*, 1997; Warner and Pierce, 1993).

The measure of offending for these illustrative analyses is the number of juveniles (ages 11 through 17) arrested for robberies in each county, pooled over the 5-year period of 1989 through 1993. The measures of the explanatory variables are based primarily on 1990 census data (United States Department of Commerce, 1992). They include (1) *residential instability*, defined as the proportion of households occupied by persons who had moved from another dwelling in the previous 5 years; (2) *ethnic heterogeneity*, computed as the index of diversity (Warner and Pierce, 1993), based on the proportion of households occupied by white versus nonwhite persons; (3) *family disruption*, indexed by female-headed households, expressed as a proportion of all households with children; (4) *poverty*, defined as the proportion of persons living below the poverty level; (5) the *unemployment* rate (coded as proportion of the workforce); (6) *proximity to metropolitan counties*, as indicated by a dummy variable with 1 being adjacent to a metropolitan statistical area and 0 being nonadjacent.[2] Also included in the analysis was the number of youth 10 to 17 years of age, which is the *population at risk* for juvenile arrests. Because states may differ in their statutes and in the organization, funding, and policies of their justice systems, it was important to eliminate from our analysis all variation between states and

[2]To ensure that single cases did not have undue influence on our results, we recoded some extreme values to values less deviant from the distribution as a whole. We set the maximum for residential stability to 0.6 (formerly 0.76; three cases recoded), that for female-headed households to 0.35 (formerly 0.42, four cases recoded), and that for unemployment to 0.12 (formerly 0.14; three cases recoded). This recoding had no substantive impact on the results, and it increases our faith in their reliability.

Table I. Descriptive Statistics

	Mean	SD
Robbery arrest rate per 100,000 per year	25.28	48.65
Population at risk in person/years	11,346	10,776
Log population at risk	8.89	1.04
Residential instability	0.39	0.07
Ethnic heterogeneity	0.26	0.20
Female-headed households	0.18	0.08
Poverty rate	0.16	0.06
Unemployment	5.64	2.60
Adjacent to metropolitan area	0.45	0.50
N of counties	264	

assess only within-state relationships pooled across the states. Therefore the model includes dummy variables representing states (with Florida serving as the omitted reference category).

3.1. The Distribution of Crime Rates

Table I presents descriptive statistics for all measures. During this 5-year period, there were 1212 arrests of juveniles for robbery in this sample of counties, which corresponds to an annual arrest rate of 40.5 per 100,000, or one arrest in 5 years for every 494 juveniles. The distribution of arrest rates is highly skewed, with zero robbery arrests of juveniles recorded in 52% of the counties, while the highest annual arrest rates were 338 and 390 per 100,000. Counties with smaller populations tended to have lower arrest rates, so the mean of robbery arrest rates across counties, 25.3, is lower than the overall arrest rate. There were zero arrests in all but one of the 47 counties with the smallest populations (700 or less). The exception is a county with two arrests in a population of 289 youths, which constitutes the seventh highest annual arrest rate. With a population this small, even a single arrest would place this county among the top 12% for arrest rate. It is clear that the data provide very crude estimates of arrest rates for any single county with a small juvenile population. Yet the lack of arrests across many small counties is strong evidence that the per capita robbery rate is lower in these counties than in counties with larger juvenile populations.

3.2. Ordinary Least-Squares Analysis

To demonstrate the purposes and use of Poisson-based regression models, I compare five analyses of the same data. The results appear in Table II. The first is an OLS regression analysis of the computed arrest rates (per 100,000 per year) for each county. The full model explains 28.4% of

Table II. Five Statistical Models of Juvenile Arrest Rates for Robbery in Nonmetropolitan Counties of Four States

	Statistical method				
Explanatory variable	OLS, rate/100,000	OLS, log(rate + 1)	OLS, log(rate + 0.2)	Basic Poisson	Negative binomial
Log population at risk					
b	11.220	0.749	1.102	1.501[a]	1.718[a]
SE	3.838	0.128	0.177	0.061	0.188
t	2.923	5.852	6.226	8.213	3.819
P	0.004	0.000	0.000	0.000	0.000
Residential instability					
b	35.573	3.017	4.366	1.567	0.162
SE	48.790	1.628	2.255	0.567	2.026
t	0.729	1.853	1.936	2.764	0.080
P	0.467	0.065	0.054	0.005	0.936
Ethnic heterogeneity					
b	63.839	2.461	3.325	2.069	2.861
SE	32.711	1.091	1.512	0.419	1.156
t	1.952	2.256	2.199	4.938	2.475
P	0.052	0.025	0.029	0.000	0.013
Female-headed households					
b	22.765	0.533	0.192	3.919	3.739
SE	71.679	2.391	3.313	1.030	2.937
t	0.318	0.223	0.058	3.805	1.273
P	0.751	0.824	0.954	0.000	0.203
Poverty rate					
b	39.474	1.405	2.181	0.499	0.021
SE	81.162	2.708	3.752	1.009	3.381
t	0.486	0.519	0.581	0.495	0.006
P	0.627	0.604	0.561	0.621	0.995
Unemployment					
b	−42.658	5.246	8.137	−1.338	0.432
SE	203.957	6.804	9.428	1.810	6.568
t	−0.209	0.771	0.863	−0.739	0.066
P	0.834	0.441	0.389	0.466	0.948
Adjacent to metropolitan area					
b	−3.944	−0.211	−0.267	−0.247	−0.458
SE	6.372	0.213	0.295	0.071	0.215
t	−0.619	−0.991	−0.905	−3.479	−2.130
P	0.537	0.322	0.365	0.000	0.034
Constant					
b	−66.020	−6.645	−11.560	−13.750	−15.243
SE	43.732	1.459	2.022	0.630	1.722
t	−1.510	−4.554	−5.717	−21.825	−8.852
P	0.132	0.000	0.000	0.000	0.000

Table II. Continued.

Explanatory variable	Statistical method: OLS, rate/100,000	OLS, log(rate + 1)	OLS, log(rate + 0.2)	Basic Poisson	Negative binomial
Model fit					
Method specific criteria					
Baseline model[b]					
MSE[c]	1853.9	2.419	4.729		α^e 1.263
R^2	0.226	0.328	0.332	0.484[f]	0.456[f]
−2LL[d]				1584.5	950.8
Full model					
MSE	1760.6	1.960	3.762		α 0.852
R^2	0.284	0.471	0.483	0.585	0.548
−2LL				1420.9	901.5
Spearman r	0.653	0.708	0.710	0.671	0.687

Note: The models also included dummy variables representing differences between the four states.
[a] t and P values computed for difference of b from 1 rather than difference of b from 0.
[b] The baseline model controls for differences between states and, in the Poisson and negative binomial models, includes log population at risk, with a fixed coefficient of 1.
[c] Mean squared error for the OLS regression models.
[d] −2 times the log likelihood for the Poisson and negative binomial models.
[e] Reflects residual variance in true crime rates, which is overdispersion beyond that expected from a simple Poisson process.
[f] See footnote 5 for a description of the computation of R^2 values for the Poisson and negative binomial analyses.

the variance in these robbery rates, which is 5.9% more than a baseline model that includes only differences between states [$F(7,253) = 2.97$, $P = 0.005$].

There are several indications that this OLS model is very poorly suited to the data. Though an arrest rate below zero would be impossible, this model yielded negative fitted values for 42 cases, and these negative values fall as much as 0.61 standard deviations below zero. Under this OLS model, the two counties with the highest arrest rates constituted extreme outliers with standardized residuals of 7.2 and 6.1, both far too large to be acceptable at any sample size. These are strong indications that the fitted values do not accurately track actual mean crime rates, so it is clear that a linear model severely distorts the relationship between these explanatory variables and county level arrest rates.

The critical assumption for the accuracy of standard errors and significance tests in OLS analysis is that the residual variance does not vary systematically across cases, and White's test for heteroscedasticity (McClendon, 1994, pp. 178–181) provides a simple and direct means of testing this assumption. This test involves an OLS regression analysis in which the squared values of the residuals serve as the dependent variables, and the fitted values of that regression will reflect mean levels of squared residuals. The independent variables in this residual analysis can be any factors suspected to be related to heterogeneity of the residuals. Because I expect that residual variance will depend on population size, but not in a linear fashion, I chose linear, squared, and cubed terms for population size as independent variables. Using absolute values of residuals rather than squared residuals as the dependent variable provided a better summary of the data. (When squared, residuals of the outliers dominated the entire sample.) This analysis indicated that the magnitude of residual variance varied widely by population size: The squared values of the fitted absolute residuals ranged from 94 to 1162 [$R^2 = 0.050$, $F(3,260) = 4.51$, $P = 0.004$].[3]

3.3. Ordinary Least-Squares Analysis of Logged Crime Rates

The most drastic shortcomings of the OLS model stem from the highly skewed distribution of arrest rates. A common strategy for addressing this problem is to transform the data so that they become less skewed. The logarithmic transformation is a common choice for this purpose because it reduces the skew and it also yields a straightforward conceptual interpretation. Under a linear model of the untransformed data, the regression coefficients indicate the difference in the mean of the dependent variable that is associated with a unit difference on the explanatory variable. After the logarithmic transformation, the regression coefficients reflect proportional differences in the mean of the dependent variable, given a 1-unit difference on the explanatory variable. For crime rates, proportional differences would appear more plausible than constant differences. We do not expect a factor that raises a crime rate from 40 per 100,000 to 60 per 100,000 to also raise a crime of 1 per 100,000 to 21 per 100,000, as would be the case for a linear model of the untransformed data. Under the proportional model produced by the logarithmic transformation, the same percentage increase would hold for both, such as 40 versus 60 and 1 versus 1.5.

The third column in Table II presents an OLS regression analysis with the natural logarithm of the arrest rate as the dependent variable. One has been added to the rates (per 100,000) before taking the logarithm because the logarithm of zero is undefined (corresponding to minus infinity). The OLS analysis of logged rates is a far better match to the data in several respects. First, in this analysis the full model accounts for 47.1% of the

[3]Strictly speaking, White's test uses the squared residual as the dependent variable and uses a χ^2 significance test due to the likely heteroscedasticity of the residuals themselves. This test would also be significant for this residual analysis [$\chi^2(3) = 13.2$, $P = 0.004$] as well as for the comparable analysis reported below [$\chi^2(3) = 28.5$, $P = 0.000$].

variance, which is a clear indication that the transformation puts the data in a form that has a closer linear correspondence to these explanatory variables. Also, in this altered metric a larger share of the explained variance is attributable to the explanatory variables rather than to differences between states [increase in $R^2 = 0.142$, $F(7,253) = 9.761$, $P < 0.001$]. Second, the range of the fitted values is not problematic under this model because negative fitted values correspond to logarithms of crime rates between zero and one. Third, the transformation also reduces problems of outliers, with the most extreme standardized residual for the OLS analysis of the transformed arrest rates now 3.2. The change in metric means that the coefficients of these first two analyses are not comparable, but the benefit of the improved correspondence between model and data is apparent in the higher t values for the three variables most strongly related to crime rates (population size, residential instability, and ethnic heterogeneity).

Though the logarithmic transformation renders the data more suitable for OLS analysis, it also makes apparent the inherent problems that require Poisson-based regression. First, rather than solving the problem of heteroscedasticity, the error variance has become even more strongly related to population size. The cubic model of the absolute residuals now acounts for 10.8% of their variance [$F(3,260) = 10.52$, $P < 0.001$], with fitted values corresponding to squared residuals that range from 0.01 to 1.98. Furthermore, the specific cases that constitute outliers also have changed. In the OLS analysis of untransformed rates, the two most extreme outliers were the counties with the highest crime rates, both of which have larger than median juvenile populations. The most extreme outlier in the OLS analysis of the transformed rates is the smallest population with any recorded arrests. The OLS assumption of homogeneity of residual variance implies that the predictive accuracy of the model is independent of population size. As we would expect from the inevitable unreliability of crime rate estimates based on small population sizes, that assumption is clearly in error.

Second, the discrete and skewed nature of crime rates for small populations presents a special problem for analyzing log transformed crime rates. Observed rates of zero will be common for small populations, in which case the transformation can be computed only after adding a constant. The common choice of adding one is highly arbitrary. The value could as easily be 1 per 1000 or 1 per 1,000,000 as the 1 per 100,000 used in the analysis just discussed, yet the choice of this constant may drastically effect the results. To see this compare Columns 3 and 4 in Table II, which differ only in that Column 3 reports an analysis resulting from adding a constant of 1 per 100,000 while the constant for Column 4 is 0.2 (corresponding to 1 arrest per 100,000 for the 5 years covered in the study, rather than for one year). This arbitrary choice results in an increase of roughly 40% in most of the regression coefficients, which is a large difference in the implied effects of these explanatory variables on mean crime rates. The reason for this change is that decreasing the constant increases the variance of the dependent variable by inflating the difference between the transformed value for the rate of zero and the transformed value for the next higher observed rate. Thus, the choice of this constant has great potential for biasing the coefficient estimates. Interestingly, changing the additive constant had minimal consequence for significance testing because standard errors grew proportionately with the coefficients, with the result that t values were essentially unchanged.

3.4. Poisson-Based Analyses

Poisson-based regression analyses successfully address the most serious problems that arise in the OLS analyses. As discussed above, Poisson-based models do not assume homogeneity of variance. Instead, residual variance is expected to be a function of the predicted number of offenses, which is in turn a function of population size. Furthermore, even though a logarithmic transformation is inherent in Poisson-based regression, observed crime rates of zero present no problem. Unlike the preceding OLS analyses of log crime rates, Poisson-based regression analyses do not require taking the logarithm of the dependent variable. Instead, estimation for these models involves computing the probability of the observed count of offenses, based on the fitted value for the mean count. As Figs. 1 and 2 demonstrate, observed rates of zero become increasingly likely as the estimated mean rate approaches zero.

The last two columns in Table II present results from a basic Poisson regression and a negative binomial regression, both estimated with the LIMDEP statistical package (Greene, 1995). Because these are maximum-likelihood estimates, likelihood-ratio significance tests can be used to determine whether more complex models provide better fit to the data than do simpler models. The test statistic is minus twice the difference between the log likelihoods for the models, and the significance level is determined by comparing this value to the χ^2 distribution with degrees of freedom equal to the number of additional parameters in the more complex model.

3.4.1. Poisson versus Negative Binomial

The negative binomial model differs from the basic Poisson by the addition of the residual variance parameter, α. The likelihood-ratio test value for the comparison between these models is 519.4 with 1 degree of freedom, indicating that the data are far more consistent with the negative binomial model than with the basic Poisson ($P < 0.001$). This result means that, on average, differences between fitted and observed crime rates are considerably larger than specified by the Poisson distribution. This overdispersion reflects some combination of unexplained variation in counties' true crime rates and positive dependence among crime events.

Comparing the fifth and sixth columns in Table II makes clear the consequence of ignoring that overdispersion. The basic Poisson model gives the impression of far greater precision in the estimated relationships than is justified. The standard errors for the basic Poisson model average only about one-third the size of the standard errors for the negative binomial, so the basic Poisson model would produce highly erroneous significance tests for the coefficients. In this example, the basic Poisson would lead us to conclude that residential instability and female-headed households are significantly related to rates of arrests of juveniles for robbery, while the negative binomial would not.

3.4.2. Fit of the Negative Binomial Model

I examined residuals from the negative binomial model to check the match between the model and data in terms of potential outliers. Standardized residuals are less useful here because the residuals are not expected to follow a normal distribution. Instead, an appropriate strategy is to use the negative binomial distribution of Eq. (4) to compute the probability of obtaining a value at least as extreme as the observed value, based on the fitted values, λ_i, and the estimate of residual variance in true crime rates, α. Because this is a tedious calculation, even using a spreadsheet program, I first computed standardized residuals[4] to identify the cases most likely to constitute outliers, and then computed probabilities for those cases. The most extreme outlier was a county with three recorded arrests. Its fitted value, λ, was 0.204, which yields a probability of 0.0043 of observing three or more arrests. Only 1 in every 233 cases should have a probability this small, but that is quite acceptable in our sample of 264. Probabilities for four cases were less than 0.02, which is no more than would be expected by chance in a sample of this size. It is especially encouraging that these four counties varied widely in population size, from the eighth to the eightieth percentile. Thus, there are good indications that the assumptions of the negative binomial model are a good match to these data and that this model

[4]Standardized residuals can be computed on the basis of the variance of the negative binomial, which is $\lambda_i + \alpha\lambda_i^2$ (Gardner *et al.*, 1995). Though the Poisson rapidly approaches the normal distribution as λ increases, this is far less true of the negative binomial with the moderate value of α obtained in this example. For instance, even with a value of λ as large as 32, standardized residuals could differ from the normal deviates corresponding to negative binomial probabilities by over 50%.

successfully addresses the confounding of population size and accuracy of crime rate estimates.

How well does the negative binomial model account for crime rates, in comparison to the OLS analyses? The R^2 values in Table II are not very helpful for this purpose because they reflect scaling differences in the dependent variable as much as differences in the success of the models. Thus, the higher R^2 values for the Poisson and negative binomial models are partly a reflection that the outcome variable in these analyses is not the crime rate, but rather the count of offenses. The Spearman rank order correlation, which is unaffected by scaling, indicates that the five models have similar success in ordering the counties by crime rate (values range from 0.66 to 0.71). When taking the metric of the outcome measure into account, we find that the negative binomial model explains substantial portions of variance in both untransformed crime rates (22.2% for negative binomial versus 28.4% for the OLS analysis of these rates) and log-transformed crime rates (22.8% for negative binomial versus 47.1% for this OLS analysis).[5] In sharp contrast, each of the OLS analyses is surprisingly unsuccessful in accounting for variance in the other metric: Fitted values from the OLS analysis of log-transformed rates account for only 8.8% of the variance in untransformed rates, while fitted values from OLS analysis of untransformed rates account for only 5.2% of the variance in log-transformed rates. It is likely that these mismatches reflect the shortcomings of both OLS approaches. A direct linear model is unsuitable for the untransformed rates, as evidenced in the impossible negative fitted values that result. Yet the OLS analysis of log-transformed rates yields fitted values that poorly match the original crime rates, despite a strong rank order correlation between the two. It appears that the arbitrary constant required for OLS analysis of log-transformed rates results in a seriously distorted metric. Poisson-based models avoid these problems and, as a result, generalize well to either response metric.

3.4.3. Results from the Negative Binomial Model

The likelihood-ratio test can be used to assess the overall contribution of the explanatory variables, comparable to the test for increase in R^2 in

[5] I computed these percentages of explained variance by transforming fitted values from each model to the metric of the observed scores, summing the squared differences between these and the observed scores, and calculating one minus the quotient of that sum divided by the total sum of squares for the observed scores. This computation corresponds to the definition of R^2 in OLS regression. Because R^2 is not part of the maximum-likelihood estimation of the basic Poisson and negative binomial models, this procedure was also used to compute the R^2 values for those models. Negative fitted values from the OLS analysis of untransformed crime rate were set to zero in order to compute the fit of that model to log-transformed crime rates. The additive constant of 1 per 100,000 was used for all log transformations of crime rates.

OLS models. For the Poisson-based analyses, the baseline model includes not only dummy variables to control for differences between states, but also the log of the population at risk with a fixed coefficient of one, as in Eq. (3). This control for population size is necessary so that the regression will be a model of per capita crime rates rather than a model of counts of crimes. For the negative binomial, the full model yields a likelihood-ratio value of 49.4 in comparison to the baseline model, which is statistically significant ($df = 7$, $P < 0.001$). Thus, we can conclude that the explanatory variables account for more variation in crime rates than would be expected by chance alone.

By the conventional 0.05 standard of statistical significance, the negative binomial analysis indicates that higher juvenile arrest rates for robbery are associated with larger populations at risk, greater ethnic heterogeneity, and being adjacent to a metropolitan area.[6] To interpret the regression coefficients for these variables, we must take into account the logarithmic transformation that intervenes between the linear model and fitted crime rates [in Eqs. (1) and (3)]. Liao (1994) explains several useful strategies for interpreting these coefficients. One relatively straightforward approach to this task follows the implication of Eqs. (1) and (3) that an increase of x in an explanatory variable will multiply the fitted mean crime rate by the $\exp(bx)$. Thus, given the coefficient of 2.861 for ethnic heterogeneity in the negative binomial model, a 10% increase in ethnic heterogeneity would multiply the rate by exp(0.286), which is 1.33. In plain English, a 10% increase in ethnic heterogeneity is associated with a 33% increase in the juvenile arrest rate for robbery. Because being adjacent to a metropolitan area is coded as a dummy variable, an increase of one in this variable corresponds to the contrast between adjacent and nonadjacent counties. Thus, the statistically significant coefficient of −0.458 indicates that counties adjacent to metropolitan areas have a 37% lower rate of robbery than those that are not because $\exp(-0.458 * 1)$ equals 0.63. [This surprising result does not replicate for analyses of other offenses reported by Osgood and Chambers (2000).]

In interpreting the results for population size, we must take into account the special role of this variable in Poisson-based analyses of aggregate rates. When the coefficient for the log of the population at risk is fixed at one [as in Eq. (3)], per capita crime rates are constant across counties with different population sizes, controlling for the other explanatory variables. The analyses reported in Table II treat that coefficient as estimated

[6]In the more extensive analyses reported by Osgood and Chambers (2000) population at risk, ethnic heterogeneity, residential instability, and female-headed households proved to be associated with most offenses, but adjacency to metropolitan areas did not.

rather than fixed, which allows for the possibility that crime rates differ with population size. In this case, however, it is necessary to subtract the value of one from this coefficient in order to determine its implications for the relationship of population size to per capita crime rates. Similarly, the statistical significance of the relationship is gauged by comparing the estimate to the value of one, rather than to zero as is the usual case.[7] A coefficient greater than one would indicate that counties with larger populations have higher per capita crime rates, while a coefficient less than one would indicate the opposite. Thus, the coefficient of 1.718 from the negative binomial analysis agrees quite closely with the value of 0.749 from the OLS analysis of the transformed crime rates. The first indicates that a doubling of the population is associated with a 64% increase in per capita robbery rates [$\exp(0.718 * \log(2)) = 1.645$], while the second implies a 68% increase [$\exp(0.749 * \log(2)) = 1.680$].

I have argued that the coefficients and significance tests based on the negative binomial (or another Poisson-based regression model that allows for overdispersion) are preferable because the other models I have reviewed rely on assumptions that are inconsistent with the data. Yet how much difference does the choice of model make? We can get some idea by comparing the coefficients and *t* values for the negative binomial analysis with those for the other analyses in Table II. Other than the OLS analysis of untransformed rates, all models specify a logarithmic relationship between fitted values and mean crime rates, so coefficients have comparable meanings across those models. In general, one would not expect an incorrect model to introduce any systematic bias, so it is surprising that estimates for many of the coefficients differ dramatically across the models. The absolute values of coefficients for residential instability, poverty rate, and unemployment are far larger in the OLS analyses than in the negative binomial analysis, while the opposite is true for female-headed households. Differences of this sort most likely are due to the role of population size in Poisson-based analyses. OLS analyses place as much weight on small counties as on large ones, but Poisson-based regression models expect error distributions in small counties to have greater variance, with the consequence that results are less influenced by small counties. This differential weighting has considerable potential for changing results in a sample such as ours, where there is a large range of population sizes.

The standard errors for the negative binomial model are most similar to those of the OLS analysis of log transformed rates, using the additive constant of one. Even here, however, standard errors for four of the seven

[7]In other words, the test statistic to be compared to the normal distribution is not the usual b/SE_b, but rather $(b-1)/\mathrm{SE}_b$.

substantive variables are at least 20% larger in the negative binomial analysis. There are far greater discrepancies in standard errors for the other models, so it is clear that significance tests may be seriously affected by applying an appropriate statistical model to aggregate data for small populations.

4. CONCLUSIONS

Using Poisson-based regression models of offense counts to analyze per capita offense rates is an important advance for research on aggregate crime data. Standard analytical approaches require that data be highly aggregated across either offense types or population units. Otherwise offense counts are too small to generate per capita rates that have appropriate distributions and sufficient accuracy to justify least-squares analysis. Poisson-based regression models give researchers an appropriate means for more fine-grained analysis. Poisson-based models are built on the assumption that the underlying data take the form of nonnegative integer counts of events. This is the case for crime rates, which are computed as offense counts divided by population size. In our example analysis of juvenile arrest rates for robbery, the Poisson-based negative binomial model provides a very good fit to the data, while OLS analyses produce outliers and require arbitrary choices that have a striking impact on results.

Poisson-based regression models free researchers to investigate a much broader range of aggregate data because they are appropriate for smaller population units and less common offenses. Yet these models are not magic. The reason they are appropriate is that they recognize the limited amount of information in small offense counts. The price one must pay in this tradeoff is that the smaller the offense counts, the larger the sample of aggregate units needed to achieve adequate statistical power. For example, this sample of 264 counties proved too small for a meaningful analysis of juvenile homicide, the least common offense examined in this study (Osgood and Chambers, 2000).

Though this article has concentrated on two of the most common Poisson-based regression models, this approach to analyzing aggregate crime rates can be implemented with virtually any of the Poisson-based regression analyses. The numerous Poisson-based models reviewed by Cameron and Trevedi (1998) offer many choices for finding a model with assumptions that best match one's data. Some models expand the range of research questions that can be addressed, such as using finite-mixture models to identify homogeneous groups of counties. Other Poisson-based models have been developed for designs with repeated measures or nested data, such as counties nested within states or multiple subpopulations nested within a sample of geographic areas. The semiparametric model of Nagin and Land (1993), which has been so influential in research on criminal careers, would be appropriate for such cases. Also, the recent version of Bryk and co-workers' (1996) HLM program implements a Poisson version of their hierarchical linear modeling approach to analyzing nested data (Bryk and Raudenbush, 1992). Thus, Poisson-based regression models should have broad applicability for the study of crime at the aggregate level.

ACKNOWLEDGMENTS

This research was supported by Grant 94-JN-CX-0005 from the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice. The author thanks Jeff Chambers for his assistance with this study, Chet Britt for comments on an early draft, and Gary Melton and Susan Limber for their support of the entire project. Points of view or opinions in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.

REFERENCES

Bailey, A. J., Sargent, J. D., Goodman, D. C., Freeman, J., and Brown, M. J. (1994). Poisoned landscapes: The epidemiology of environmental lead exposure in Massachusetts. *Soc. Sci. Med.* 39: 757–766.

Bryk, A. S., and Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage, Newbury Park, CA.

Bryk, A. S., Raudenbush, S. W., and Congdon, R. (1996). *HLM: Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs*, Scientific Software International, Chicago.

Cameron, A. C., and Trivedi, P. K. (1998). *Regression Analysis of Count Data*, Cambridge University Press, Cambridge.

Gardner, W., Mulvey, E. P., and Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial. *Psychol. Bull.* 118: 392–405.

Greenberg, D. F. (1991). Modeling criminal careers. *Criminology* 29: 17–46.

Greene, W. H. (1995). *LIMDEP: Version 7.0 Users Manual*, Econometric Software, Plainview, NY.

King, G. (1989). *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*, Cambridge University Press, Cambridge.

Liao, T. F. (1994). *Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models*, Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–101, Sage, Newbury Park, CA.

Maltz, M. D. (1994). Operations research in studying crime and justice: Its history and accomplishments. In Pollock, S. M., Rothkopf, M. H., and Barnett, A. (eds.), *Operations Research and the Public Sector, Volume 6 of Handbooks in Operations Research and Management Science,* North-Holland, Amsterdam, pp. 200–262.

McClendon, McK. J. (1994). *Multiple Regression and Causal Analysis*, F. E. Peacock, Itasca, IL.

McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed., Chapman and Hall, London.

Nagin, D. S., and Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model. *Criminology* 31: 327–362.

Osgood, D. W., and Chambers, J. M. (2000). Social disorganization outside the metropolis: An analysis of rural youth violence. *Criminology* 38: 81–115.

Rowe, D. C., Osgood, D. W., and Nicewander, W. A. (1990). A latent trait approach to unifying criminal careers. *Criminology* 28: 237–270.

Sampson, R. J., Raudenbush, S. W., and Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science* 177: 918–924.

United States Department of Commerce (1992). Summary Tape Files 1 and 3, 1990 Census.

Warner, B. D., and Pierce, G. L. (1993). Reexamining social disorganization theory using calls to the police as a measure of crime. *Criminology* 31: 493–517.