Statistical Models for Political Science Event Counts: Bias in Conventional Procedures and Evidence for the Exponential Poisson Regression Model
Author(s): Gary King
Source: *American Journal of Political Science,* Vol. 32, No. 3 (Aug., 1988), pp. 838-863
Published by: Midwest Political Science Association
Stable URL: https://www.jstor.org/stable/2111248
Accessed: 08-01-2020 16:32 UTC

Statistical Models for Political Science Event Counts: Bias in Conventional Procedures and Evidence for the Exponential Poisson Regression Model*

Gary King, *Harvard University*

This paper presents analytical, Monte Carlo, and empirical evidence on models for event count data. Event counts are dependent variables that measure the number of times some event occurs. Counts of international events are probably the most common, but numerous examples exist in every empirical field of the discipline. The results of the analysis below strongly suggest that the way event counts have been analyzed in hundreds of important political science studies have produced statistically and substantively unreliable results. Misspecification, inefficiency, bias, inconsistency, insufficiency, and other problems result from the unknowing application of two common methods that are without theoretical justification or empirical utility in this type of data. I show that the exponential Poisson regression (EPR) model provides analytically, in large samples, and empirically, in small, finite samples, a far superior model and optimal estimator. I also demonstrate the advantage of this methodology in an application to nineteenth-century party switching in the U.S. Congress. Its use by political scientists is strongly encouraged.

1. Introduction

This study is concerned with statistical models for event count data. Event counts are variables that have for observation i ($i = 1, \ldots, N$) the number of occurrences of an event in a fixed domain. The domain for each observation may be time—as in a month, year, hour, or some appropriate interval—or space—as in a geographic unit, an individual, or others.

Dependent variables of this type exist in every major journal and empirical field in the discipline, often representing central concepts or concerns. The largest number of event counts is probably in international relations, where massive databases record the number of actions each nation or political group takes with respect to another (e.g., Azar and Sloan, 1975). But examples from other fields abound: the number of presidential vetoes per year (Rohde and Simon, 1985); the number of congressional staff members engaged in casework services and the number of trips to the home district for each member of the House of Representatives (McAdams and Johannes, 1985); the number of seats the president's party lost in each midterm congressional election (Campbell, 1985); the size of Medicaid caseloads (Hanson, 1984); the number of persons recorded in LBJ's daily diary as having been present at a particular White House meeting (Sigelman and McNeil, 1980); the number of months that a parliamentary cabinet endures

(Robertson, 1984); the number of members of the House and Senate who switch political parties each year (King and Benjamin, 1985); the number of citizen-initiated and support-related political activities engaged in and reported by Soviet émigrés (Di Franceisco and Gitelman, 1984); the number of coups d'etat per year for black African states (Johnson, Slater, and McGowan, 1984). There are many other examples.

In the sections that follow, I show the data generation process of event counts to be Poisson (sec. 2). I then introduce the exponential Poisson regression (EPR) model as the appropriate method, directly deducible from the data generation process (sec. 3). The next section shows that the two models used most frequently in political science for analyzing this type of data are either misspecified (the OLS model in sec. 4) or biased and inconsistent (the logged OLS model in sec. 5). Since statistical theory is known only for unrealistically large sample sizes, I use Monte Carlo experiments to demonstrate the empirical unbiasedness of the EPR model in finite samples and the bias and inefficiency of the logged OLS model of event counts even in very large samples (sec. 6). A brief analysis of an empirical example (sec. 7) and concluding remarks (sec. 8) are also provided. Appendix A provides details of the proof used in section 3. Appendix B reviews readily available computer programs that can be used to estimate the EPR model.

2. The Data Generation Process of One Event Count Observation

The data generation process usually assumed to produce event counts is Poisson.[1] This process arises naturally in many situations commonly analyzed by political scientists. This section presents the substance of a proof that all event count data that meet a few modest assumptions arise from a Poisson process (mathematical details appear in Appendix A).

Consider a model for what will be *one observation* in the Poisson regression model to be discussed in later sections. Let y_t be the number of recorded events (a nonnegative integer) at an instant in time, t; t does not refer to the observation number, but refers to the time that has passed in recording events within this one observation; y_t never decreases, and at certain instances in time, the number of events increases by one. This process is not observed until the end of the observation period, when the total number of events occurring within the period are recorded.

---

[1] Next to the normal, and certainly among discrete distributions, the Poisson probability distribution is often considered the most important. Interestingly, although he was a mathematician, Simeon Denis Poisson's (1837) invention was in the context of a "political science" work on the application of mathematical probability to judicial administration (see Haight, 1967, p. 113). Furthermore, the methodological problems and opportunities presented by event count data are focused in political science. Economics and psychology have occasional applications, and while sociology has a reasonable number, it does not approximate the many myriad applications in political science. See Haight (1967) for applications in the natural and physical sciences.

A mathematical model emerges by assuming that these events occur at "random" within this one observation with a finite rate of occurrence, θ ($0 < \theta < \infty$). "Random" is defined in the Markov sense by assuming that in a very short interval of time, Δt, the probability of one additional count is independent of past and present numbers of events. This assumption basically implies that during the observation period the expected rate of occurrence of the next event either remains constant (and equal to θ), or at least does not change in response to the number of observed events, and that the random error around θ at one instant in time is uncorrelated with the random error at the next point in time. This rate of occurrence may and usually does change across observations, but within this one observation it is assumed fixed or at least unresponsive to the number of observed events. This is a relatively weak requirement that should hold in most political situations. When it does not, θ refers to the average rate within the observation; in this case, event count data would mean the loss of some information and "event history" methods (Allison, 1984) would be preferable. However, political data for both independent and dependent variables are rarely available in this detailed form. Thus, "event-count analysis deserves further study" (Tuma and Hannan, 1979). Indeed, even this assumption can be relaxed, since the aggregation "of a large number of independent and uniformly sparse [variables] *of any type* is approximately a Poisson process" (Amburgey and Carroll, 1984, pp. 41–42). For continuous variables in ordinary regression analysis, the normal distribution is often similarly justified as the sum of many unmeasured variables. From these assumptions—y_t being a random event count with a finite, positive, and constant (or "unresponsive") mean—one can derive the Poisson probability distribution (see Figure 1 and Appendix A).

Thus, we have constructed a Poisson distribution for one observation *directly* from an underlying mathematical model built on a few plausible assumptions about the political phenomenon generating the counts. The Poisson distribution also arises *indirectly* from the limiting form of a very large number of other distributions and naturally occurring situations (Haight, 1967; Johnson and Kotz, 1969).[2] I believe that the case made here for typical event count data being generated by a Poisson process is stronger than the case usually made for most other political science variables being distributed normally. Since the mathematical process underlying event count data is relatively clear and very common, these data are some of the best in the discipline. Certainly when using these data, political scientists ought to more fully exploit them and their special properties.

---

[2] The negative binomial distribution is one that tends to the Poisson, but it can also arise when the Poisson parameter is allowed to vary according to a Gamma distribution, possibly representing unobserved heterogeneity across observations. If the data are negative binomial but the Poisson model is used, estimates will still be consistent (and usually very similar), but standard errors will be deflated somewhat (see Gourieroux, Monfort, and Trognon, 1984; King, 1987c).

3. The Poisson Regression Model

I move now to the more usual situation where there are N event count observations. In order to shift attention from within one observation (the dynamics of which are not observed) to across observations, I drop the subscript t on y. I assume that the time interval or domain size characterizing each observation is of the same length or size, although this assumption can be dropped. Then let y_i $(i = 1, \ldots , N)$ be a set of Poisson processes representing the number of events recorded for each observation i. Note that the unit of analysis may be time or space or both. Thus, the number of violent acts per day in the United States (say 1950–80) and the number of violent acts per state (say on 5 July 1979) in the United States are equally valid examples. Of course, these two data sets are probably useful for answering different substantive questions, but either can be analyzed using this model. For simplicity, I assume that the absence of autocorrelation, that is, $C(y_i, y_j) = 0$ for all $i \neq j$, although this assumption can also be relaxed. Since the mean of the Poisson distribution at each observation θ_i is equal to its variance, heteroskedasticity is guaranteed for all nontrivial examples. In the probability distribution, larger values of θ_i have a larger disturbance variance. This is plausible, since as the mean gets closer to its lower bound of zero, the variance must be smaller. Note how, in Figure 1, the variance is larger for $\theta_i = 2.0$ than for $\theta_i = 0.75$. Also, due to the positively skewed Poisson distribution, the heteroskedasticity is strongly asymmetric, since it is not bounded on the right (for presentation purposes, Figure 1 stops where the probability of a larger number of events is very close to zero). The appropriate estimation technique to be discussed below takes full advantage of this additional information.

To fruitfully analyze these data, let the mean of y_i be a function of a vector of explanatory variables:
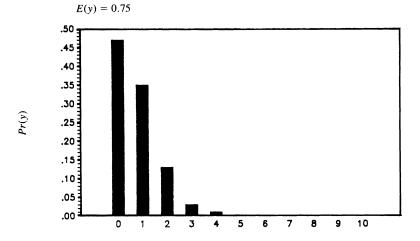
$$E(y_i \mid X) \equiv \theta_i = f(x_i' \beta) \quad \text{or} \quad y_i = f(x_i' \beta) + \varepsilon_i \tag{3.1}$$

where,

$$y_i \sim \text{Poisson}(\theta_i) \qquad \text{or} \qquad \varepsilon_i \sim \text{Poisson}(\theta_i) - \theta_i$$

To complete this specification, only $f(.)$ need be chosen. If $f(.)$ were an identity function, then equation 3.1 would be $y_i = x_i' \beta + \varepsilon_i$, and β could be estimated by an ordinary least squares (OLS) regression analysis. Although the Poisson distribution of y_i would make this procedure quite inefficient, the estimator would be an unbiased estimate of the linear approximation to the conditional expectation function. The additional problem with OLS, however, is that it does not constrain the expected number of events θ_i to be positive, as it must be. We therefore limit $f(.)$ to the class of positive valued functions. To narrow this to a uniquely appropriate function, consider what the effect of x on y ($\partial y / \partial x$) should be. We have

FIGURE 1

The Poisson Probability Distribution

$E(y) = 0.75$

$E(y) = 2.00$

already established that this effect should not be β, since that would imply a linear function and would leave open the possibility of the expected number of events being less than zero. I propose that the most reasonable effect is $\beta\theta_i$, so that a fixed change in x would have a greater effect on y_i if the expected value were larger. In other words, this specification indicates that the "effort" (in terms of a

change in x) that it would take to move y_i from 0 to 1 should be proportionally greater than the effort required to move y_i from, say, 20 to 21.

For example, suppose we wanted to know the effect of a congressional representative's activity in Washington (e.g., bills introduced or participation on roll calls) on national press coverage (e.g., the number of mentions in the *New York Times;* see Cook, 1986). Consider two members of the House, one of whom routinely receives considerable press coverage (a high initial expected coverage) and another who receives very little (a low initial expected coverage). A linear form implies that, if both representatives increased their activity level by the same amount in a particular year, their expected levels of coverage (θ) during that year would each increase by the same number of mentions (β). More plausibly, however, the member with the higher *initial* level of expected coverage is likely to translate the same amount of additional activity into proportionately more coverage than the other member. This idea can be incorporated as follows: let $\beta\theta_i$ be the increase in coverage that results from a unit increase in activity. Then, the greater the expected number of mentions, the larger will be the effect of activity on coverage.
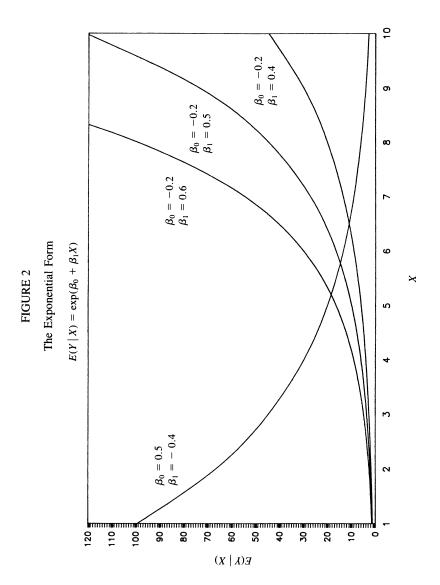
If $\beta\theta_i$ is the appropriate effect, then the functional form is $E(y_i \mid X) = \exp(x'\beta)$ (using the fact that exponentiation is the only function that is its own derivative). Although there are other functions that would restrict the expected value to be positive, there are few that *also* have very plausible effects. By simultaneously applying both of these criteria, I am led to this particular function. Exponentiation also happens to be the function most commonly used with the Poisson distribution in other contexts and is implied in the very widely used log-linear models of contingency table counts (Plackett, 1981). Note that θ_i need not be an integer and may be taken to represent either the literal mean population value for observation i, since the average of a set of integers may be a real number, or the expected number of events measured more precisely. Figure 2 portrays several examples of how this curve gets steeper as $\theta_i \equiv E(y_i \mid X)$ gets larger. It also makes apparent how the zero bound on the predicted number of events "bends" the curve up from the bottom.

Taking into account the heteroskedasticity and the other features of the model indicated or implied above, the method of maximum likelihood can be used to derive estimates of β.[3] See Appendix B for information on computer pro-

---

[3] First assign t in equation A.4 to have a constant value of 1. Then the joint likelihood function of $y_1, \ldots, y_n$ is:

$$L \equiv g(y_1, \ldots, y_n; \theta_1, \ldots, \theta_n) = \prod_{i=1}^{n} \frac{e^{-\theta_i}\theta_i^{y_i}}{y_i!} \tag{3.2}$$

where $\theta_i = \exp(x_i'\beta)$ for $i = 1, \ldots, n$. The log-likelihood function is then equal to $q \equiv \sum_{i=1}^{n} \ln(L_i)$, where

FIGURE 2

The Exponential Form

$$E(Y \mid X) = \exp(\beta_0 + \beta_1 X)$$

grams that will estimate this model. When y_i (for all or nearly all observations) is "large," it would be possible to analyze this sort of data by linear least squares techniques. The reason is that the Poisson variable y_i becomes approximately normal when the number of counts become large. (Note that the number of observations has no effect on this, only the number of events within each observation.) How large must the counts be? Every situation is different, but it seems reasonable to apply the well-known rule of thumb for the central limit theorem: if the number of events counted for each (and almost every) observation is greater than about 30, then it is probably safe to assume that the disturbances are approximately normally (or log-normally) distributed.[4] Since there is no problem created by using this method with larger counts, the exponential Poisson regression model is the safer solution with all event count variables.

Unfortunately, although the exponential Poisson regression (EPR) model is most appropriate for even count data, it has only very rarely been applied either in political science or anywhere else. See Maddala (1983, pp. 51–54), Jorgenson (1961), and Amburgey and Carroll (1984) for differing theoretical treatments of this model. See King (1987a) for a political application. One problem with this previous research is that the beneficial properties of this estimator have been established only because it is based on the method of maximum likelihood. This means that only asymptotic (large N) results are known. Thus, section 6 presents Monte Carlo experiments to assess the model in small, finite samples. An example is also provided in section 7. I move now to the two more common methods of analyzing these data.

4. Ordinary Least Squares and Event Counts

By far the most common model applied in political science to the event count data described in sections 1 and 2 is that implied by ordinary least squares (OLS) regression analysis. OLS provides an unbiased linear estimator, and this is unaffected by different distributional assumptions. However, there are several serious problems in using event count data with the OLS model.

First, OLS assumes a linear relationship, $E(y \mid X) = X\beta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots$. This is an implausible functional form for two reasons: (1) it often results in predicted event counts that are less than zero and therefore meaning-

---

$$\ln(L_i) = -e^{x_i'\beta} + (x_i'\beta)y_i - \ln(y_i!) \tag{3.3}$$

In estimation, $\ln(y_i!)$ may be dropped, since it does not vary with β. Because there is no analytical solution for b that maximizes L, numerical maximization methods must be used (see Maddala, 1977, pp. 176–81). Using Berndt et al.'s (1974) modified method of scoring in practice, I find that even large problems converge relatively quickly.

[4] There are also weighted least squares (WLS) estimators for this problem that are analogous to those championed by Kritzer (1978) for the analysis of contingency tables (El-Sayyad, 1973; McCullagh and Nelder, 1983).

less. Moreover, a "truncated linear" model, where negative fitted values are forced to zero, makes unrealistic assumptions at and near the cutoff point. Furthermore, (2) it makes the unrealistic assumption that the difference between zero and one event occurring in a particular time interval is the same as the difference between, say, 20 and 21 events. Thus, the true relationship is not linear, and a linear approximation would not in most cases even be a reasonable working assumption. OLS is an unbiased estimator of a linear conditional expectation function (CEF); the problem here is that the CEF is neither linear nor necessarily close to linear.

Second, the statistical inefficiency (the variance of the estimates across samples) of the OLS estimator is much higher than it could be. By taking into account neither the heteroskedasticity, the particular asymmetric form of the heteroskedasticity, the correct functional form, nor the underlying Poisson distribution of the disturbances, OLS does not use all available information in the estimation. Insufficiency and inefficiency result.

These statistical problems are more than just technical points. They usually result in substantively biased conclusions. In applications, coefficients will have the wrong size and will often have the incorrect sign. Questions such as "How many disruptive events will occur next month if unemployment decreases to 10 percent?" will many times yield nonsensical answers like, "The estimates indicate that there will be about negative four disruptive events." (This does not mean that there will be four less events; it means that the predicted number of events is -4.0.) Furthermore, estimates will often be very imprecise, making many empirical analyses inconclusive. In fact, since the standard errors and test statistics are themselves biased, there will usually be no indication of this imprecision. Unfortunately, these serious criticisms of the OLS model apply to most existing analyses of event count data in political science.

5. The Logged OLS Model of Event Counts

As an attempt at a partial solution to the first set of problems with the OLS model of event counts, some political scientists have regressed the natural log of y on X. This is a seemingly plausible specification, since logging y changes the linear model by discounting large values of y. This is believed to meet the functional form objection of the OLS model of event counts.

The problem is that the logged OLS (LOLS) model adds additional complications. The primary objection is that the conditional expectation of $\ln(y_i)$, given X and $y_i \sim \text{Poisson}(\theta_i)$, is approximately negative infinity, even in finite samples.[5] Since this conditional expectation is not at all meaningful, the es-

---

[5] Recalling that $\ln(0) \approx -\infty$, $E[\ln(y_i) \mid X] = \sum_{y_i=0}^{\infty} \ln(y_i) Pr(y_i) = -\infty$. This problem is not solved by using a special Poisson distribution that does not include the zero class, since, although the CEF of $\ln(y_i)$ given X for this distribution exists (Johnson and Kotz, 1969, pp. 104–09), it omits these sometimes critical observations.

timator is biased, inefficient, insufficient, inconsistent, and theoretically without merit. Letting b^{LOLS} be the LOLS estimator,

$$E[b^{LOLS} \mid X] = (X'X)^{-1} X' E[\ln(y_i) \mid X] \approx -\infty. \tag{5.1}$$

This problem is often "solved" by regressing $\ln(y_i + c)$ on X, where c is some arbitrary small positive constant (usually 0.01) added to y_i so that logs can be taken.[6] This procedure makes the expected value in equation 5.1 finite, but it cannot be expressed in closed form and is still biased. Indeed, as will be seen, the small constant can create even more serious problems.

To observe how this LOLS model provides a certain approximation to the true model, recall that $E(y \mid X) \equiv \theta_i = \exp(x_i\beta)$ implies $y_i = \theta_i + \varepsilon_i$, which in turn can be expressed as

$$\ln(y_i + c) = \ln(\theta_i + c + \varepsilon_i) \tag{5.2}$$

This is the closest that the EPR specification will ever get to the LOLS specification, since the log of a sum is *not* the sum of the logs.[7] The slope coefficient of the best linear approximation to equation 5.2 can now be found by taking its first derivative:

$$\frac{\partial \ln(y_i + c)}{\partial x} = \beta \left[\frac{\theta_i}{y_i + c} \right] \tag{5.3}$$

The LOLS coefficient is a sample estimate of the right hand side of 5.3, where β in that equation is the true coefficient (estimated correctly by EPR). The "bias" is then the term in parentheses; y is a nonnegative integer; and θ and c are positive. This means that the coefficient from the LOLS approximation generally has the same sign as the more appropriate EPR estimation. Furthermore, since the distribution of y_i around θ_i is strongly asymmetric (with the longer tail pointing away from zero), it is very often true that θ_i is greater than y_i. To get a feel for why this is true, consider the extreme case of $y_i = 0$. Since θ_i is always greater than zero, in this case it will be always greater than y_i. Indeed, in most empirical applications with several zero observations, the sum of the residuals ($e_i = y_i - \exp(X_i b)$) may be negative, albeit only very slightly so. This is plausible, since

[6] In their review of this model applied to contingency table counts, Bishop, Fienberg, and Holland (1975, p. 354) point out that "the effects of replacing such zero values for observed cell counts by some positive value . . . have not been definitively investigated." The results below should therefore have application to models of contingency table counts. In another context and for a different purpose, Plackett (1981, p. 5) suggests that 0.5 might be the logical value to add to y, since $E[\ln(y + 0.5)] - \ln(\theta)$ tends toward 0 as θ tends toward infinity. The problem with this logic is that the advantage of the EPR model, and the need for the small constant to be added, are greatest when θ is smaller.

[7] If the specification were changed to include a multiplicative disturbance term, logs could be taken, but a similar bias discussed above, and analyzed below, still remains.

whenever the rate of occurrence is positive but very small, one would expect many of the realized values to be zero.

Thus, when the c is relatively small (it must be greater than zero), the LOLS approximation is likely to provide estimates that are too large in absolute value. If a larger value of c is chosen, however, an incrementally smaller value for the LOLS estimates will result; in these cases, it is also possible for the LOLS results to be too small. Furthermore, there is no optimal choice for c across different empirical examples. Thus, the "arbitrary" choice of a small constant is not arbitrary at all in this case; this choice can drastically change one's substantive conclusions. In the next section, the most common value for c (0.01) is used in Monte Carlo experiments to explore the actual empirical consequences of using the LOLS estimation. Then, in section 7, a political science example is fit in several ways with differing values for c to demonstrate this result.

It pays to remember that the LOLS estimator is biased whether or not a small constant needs to be added. Furthermore, the results in any one empirical application will not have any necessary relationship with the true parameter values. The estimates may be too small, too large, or even have different signs. Thus, logging y_i and running OLS is not only dangerous, but it is an inappropriate solution to this problem. The lack of any meaningful underlying model makes this estimation useless and a linear approximation to it quite arbitrary.

Another method, only infrequently used, is to take the square root of y and regress it on x. At first, this seems appropriate since (1) the square root of zero is defined and no constant need be added, and (2) the square root operation approximately stabilizes the variance of a Poisson variable, thus avoiding the problem of heteroskedasticity (actually, $[y + 3/8]^{1/2}$ does somewhat better; see Haight, 1967, p. 76). However, the resulting variable is still skewed and is not approximately normal; thus, OLS will still be inefficient. (To approximate normality, one could use a different transformation, $y^{2/3}$, but this would reintroduce heteroskedasticity.) Furthermore, when expressing $y_i^{1/2} = X\beta + \varepsilon$ in terms of the original event count variable, the model does not appear substantively interesting. This can be seen by letting $\omega \equiv X\beta$; then, the closest one can come to a reasonable model is $y_i = \omega_i^2 + 2\omega_i\varepsilon_i + \varepsilon_i^2$, or $E(y_i) = \omega_i^2 + E(\varepsilon_i^2)$, where $\varepsilon_i = y_i^{1/2} - \omega_i$. This implausible form has the mean event count as a function of X and the variance of the disturbances, both of which vary over observations. Attempting to reduce the last expected value calculation makes the problem with this model even more apparent, since letting $E(y) = \lambda$ yields the awkward and substantively uninteresting formula:

$$E(y^{1/2}) \approx \lambda^{1/2} - \frac{1}{8\lambda^{1/2}} - \frac{7}{128\lambda^{3/8}}$$

A related, but somewhat more reasonable, model is $E(y_i) = \omega_i^2$ or $y_i = \omega_i^2 + \varepsilon_i$, where $\varepsilon_i = y - \omega_i^2$. This model is reasonable, since the effect is similar to—

although not as direct or useful as—the EPR model: $\partial E(y)/(\partial X) = 2\omega_i\beta$. This could be estimated with maximum likelihood in a manner similar to the EPR model. The coefficients could be interpreted as if they were approximately half of the EPR coefficients. But there is no estimation advantage to this model, since applying OLS to this model would not be appropriate. The linear approximation to $y^{1/2} = (\omega_i^2 + \varepsilon)^{1/2}$ is not β:

$$\frac{\partial y^{1/2}}{\partial x} = \beta \left[\frac{\omega}{(\omega^2 + \varepsilon)^{1/2}} \right]$$

The term in square brackets represents the OLS bias as a linear approximation to this nonlinear model. Unfortunately, even the size and the sign of this bias are indeterminate, so the situation may even be worse than with the LOLS model.

6. Monte Carlo Experiments

As suggested above, two important reasons lead us to proceed with Monte Carlo experiments given the above analytical results. First, the EPR model is based on asymptotic theory. Maximum likelihood estimators such as this are normally distributed, are consistent, have variances equal to the Cramer-Rao lower bound (i.e., minimum variance), and have other desirable properties, but all only in infinite samples. In asymptotic theory, the EPR model is universally better than the OLS or LOLS models of event counts. However, in the sometimes small, but always finite, samples commonly used in political science data analyses, this statistical theory may or may not apply. For example, Amburgey and Carroll (1984, p. 53) conclude their work by arguing that "Monte Carlo studies are needed to determine how much is gained by [these] methods . . . especially in small samples."

Second, although sections 4 and 5 have analytically shown the very serious problems with the most commonly used event count models in political science, no argument exists attempting to explain in empirical terms just what so many computer runs have produced. It is possible, although not very probable, that empirical applications of the LOLS event count model might not be too biased or inefficient.

Monte Carlo experiments were therefore conducted according to the following procedures (see Hendry, 1984): β and X were arbitrarily chosen. A (3×1) β vector was arbitrarily chosen to include $\beta_1 = 2.0$, $\beta_2 = 0.4$, and $\beta_3 = -3.0$. Although not presented below, many other values in the parameter space were also tried, with results quite typical of those presented here. The matrix of exogenous factors X has a constant term and two "variables." As a result of these choices, the vector of expected values $E(y) = \theta = \exp(X\beta)$ had an empirical minimum (for 2,000 observations) of 0.2063 and maximum of 10.8228.

Experiments were conducted at 12 different sample sizes, including $n = 10$, 20, 30, 40, 50, 100, 150, 200, 500, 750, 1,000, and 2,000. For each sample

Gary King

TABLE 1

Bias and Inconsistency

Sample Size	Estimator	$\beta_1 = 2$	$\beta_2 = 0.4$	$\beta_3 = -3$	Sum of Squared Error
10	EPR	2.0009	0.3465	−3.2372	0.05913
	LOLS	2.0064	0.4329	−4.8140	3.29172
20	EPR	1.9879	0.4454	−3.0771	0.00815
	LOLS	2.0626	0.6074	−5.0461	4.23346
30	EPR	2.0289	0.3650	−3.1173	0.01582
	LOLS	2.0411	0.6799	−5.2344	5.07258
40	EPR	2.0341	0.3046	−2.9643	0.01154
	LOLS	2.0569	0.7019	−5.0925	4.47294
50	EPR	1.9623	0.4656	−3.0205	0.00614
	LOLS	2.0770	0.6919	−5.1575	4.74594
100	EPR	2.0341	0.3851	−3.0731	0.00673
	LOLS	2.0878	0.6630	−5.2330	5.06317
150	EPR	1.9864	0.4084	−3.0045	0.00028
	LOLS	2.1601	0.5724	−5.2589	5.15798
200	EPR	2.0186	0.3858	−3.0365	0.00188
	LOLS	2.1597	0.6321	−5.2683	5.22456
500	EPR	2.0032	0.3932	−2.9899	0.00016
	LOLS	2.1602	0.6217	−5.2170	4.98990
750	EPR	1.9931	0.3986	−2.9870	0.00022
	LOLS	2.1411	0.6314	−5.2184	4.99475
1,000	EPR	2.0002	0.4009	−2.9971	0.00001
	LOLS	2.1465	0.6179	−5.2142	4.97162
2,000	EPR	1.9986	0.3987	−2.9960	0.00002
	LOLS	2.1465	0.6266	−5.2248	5.02254

NOTE: Each cell entry is an average of 100 independent random samples for a fixed sample size and estimator.

size, 100 EPR and LOLS models were estimated.[8] Across these trials, X and β remained fixed. y_i ($i = 1, \ldots, N$) was generated for each trial to reflect the underlying process: a random draw from a Poisson distribution with parameter θ_i

[8]The OLS model was not estimated in these experiments, since it cannot be plausibly compared with the EPR model. If in the EPR model we let $\theta_i = X\beta$, the OLS model could be compared. However, this would not be very interesting, since in this case even the EPR model could (and probably would) produce negative fitted values and other undesirable and implausible results. In addition, the LOLS model is also a least squares technique, and most results from these experiments apply analogously to the OLS model. In empirical applications, my experience conforms to the results presented below. The most commonly used constant, $c = 0.01$, was added to y_i to permit logs to be taken in this LOLS model.

$(i = 1, \ldots, N)$. The results of these experiments are evaluated in three stages: bias and inconsistency (Table 1); efficiency and standard error bias (Table 2); and the sampling distribution of the EPR coefficients (Table 3).

An estimator is said to be "unbiased" if the average value of the parameter estimates across an infinite number of samples drawn from the same population is equal to the respective population parameter. As an approximation, Table 1 presents the average value of the estimated coefficients across 100 samples (for each sample size and estimation procedure). If taking an infinite number of samples were possible, an estimator would be unbiased when the averages reported in the table of the estimate of β_1 would equal 2, β_2 would equal 0.4, and β_3 would equal -3.

All the rows marked EPR in Table 1 have values quite close to 2, 0.4, and -3, respectively, indicating unbiasedness. However, the rows marked LOLS differ substantially from the population parameter, even for very large sample sizes. A closer look at the individual coefficients indicates that they are, in fact, systematically biased. For every average estimate, and for every sample size, the LOLS results are all larger in absolute value than their respective parameters and than the EPR estimates.[9] The estimate of the third parameter ($\beta_3 = -3$) is especially noteworthy. Even at $n = 2,000$, while the EPR coefficient is only 0.1 percent off the -3.0 mark, the LOLS coefficient is more than 74 percent larger than it should be. Thus, the EPR estimator appears unbiased, whereas the LOLS estimator appears biased away from zero.

An estimator is said to be "consistent" if the sampling distribution collapses to a spike over the population parameter as the sample size goes to infinity. Of course, it is not possible to take a sample size of infinity, but an estimator that is consistent should generally improve (produce estimates closer to the population values) as the sample size increases. In the case of Table 1, the average estimates should generally get closer to the target values of 2, 0.4, and -3 as the sample size increases.

The EPR estimator does appear consistent; the average of the 100 samples generally gets closer to the population parameters as the number of observations increases. However, the LOLS estimator does not move closer to the target parameter values as larger sample sizes are used.

These results can also be seen in the last column of Table 1. This column is a convenient summary of the amount of error in the set of three estimate averages in each row. If b_i is the estimate, the number in that column is $\sum_{i=1}^{3}(b_i - \beta_i)^2$ [e.g., in the first row $0.05913 = (2.0009 - 2)^2 + (0.3465 - 0.4)^2 + (-3.2372 - 3)^2$]. If the estimates are exactly correct, this number is zero; as the estimates diverge further from the population parameters, this number becomes larger. The

---

[9]Exactly half of the EPR estimates are larger than the parameters and half are smaller. For those cases in which the EPR estimates were larger than the parameters, the LOLS estimates were all larger still.

TABLE 2

Efficiency and Standard Error Bias

Sample Size	Estimator	$\beta_1 = 2$	$\beta_2 = 0.4$	$\beta_3 = -3$
10	EPR S.E.	0.6416	1.3501	1.3426
	$\text{Std}(b^{\text{EPR}})$	0.4081	0.8684	0.8040
	$\text{Std}(b^{\text{LOLS}})$	1.0457	2.3126	1.4785
20	EPR S.E.	0.3962	0.6293	0.6655
	$\text{Std}(b^{\text{EPR}})$	0.2659	0.3883	0.5754
	$\text{Std}(b^{\text{LOLS}})$	0.7567	1.3750	1.0020
30	EPR S.E.	0.3321	0.5093	0.5203
	$\text{Std}(b^{\text{EPR}})$	0.2659	0.4016	0.4762
	$\text{Std}(b^{\text{LOLS}})$	0.6771	1.1085	0.8210
40	EPR S.E.	0.2664	0.4174	0.4138
	$\text{Std}(b^{\text{EPR}})$	0.2489	0.3846	0.3757
	$\text{Std}(b^{\text{LOLS}})$	0.6571	1.1511	0.6925
50	EPR S.E.	0.2458	0.4120	0.3698
	$\text{Std}(b^{\text{EPR}})$	0.1925	0.3217	0.3559
	$\text{Std}(b^{\text{LOLS}})$	0.5170	0.9509	0.7019
100	EPR S.E.	0.1688	0.2554	0.2480
	$\text{Std}(b^{\text{EPR}})$	0.1350	0.1978	0.2358
	$\text{Std}(b^{\text{LOLS}})$	0.4396	0.7451	0.4499
150	EPR S.E.	0.1354	0.2041	0.2044
	$\text{Std}(b^{\text{EPR}})$	0.1306	0.2032	0.2085
	$\text{Std}(b^{\text{LOLS}})$	0.3387	0.5817	0.4355
200	EPR S.E.	0.1165	0.1638	0.1776
	$\text{Std}(b^{\text{EPR}})$	0.0948	0.1404	0.1602
	$\text{Std}(b^{\text{LOLS}})$	0.2832	0.4472	0.3122
500	EPR S.E.	0.0786	0.1069	0.1122
	$\text{Std}(b^{\text{EPR}})$	0.0745	0.0976	0.1026
	$\text{Std}(b^{\text{LOLS}})$	0.2340	0.3281	0.2231
750	EPR S.E.	0.0612	0.0846	0.0909
	$\text{Std}(b^{\text{EPR}})$	0.0588	0.0772	0.0953
	$\text{Std}(b^{\text{LOLS}})$	0.1744	0.2588	0.1758
1,000	EPR S.E.	0.0511	0.0722	0.0768
	$\text{Std}(b^{\text{EPR}})$	0.0510	0.0681	0.0778
	$\text{Std}(b^{\text{LOLS}})$	0.1501	0.2057	0.1658
2,000	EPR S.E.	0.0350	0.0504	0.0551
	$\text{Std}(b^{\text{EPR}})$	0.0354	0.0457	0.0556
	$\text{Std}(b^{\text{LOLS}})$	0.0854	0.1232	0.1086

NOTE: S.E. is the average standard error over 100 trials; $\text{Std}(b^{\text{EPR}})$ is the standard deviation of the EPR coefficients across trials; $\text{Std}(b^{\text{LOLS}})$ is the standard deviation of the LOLS coefficients across trials.

LOLS error is always larger than that for the EPR estimator. In fact, the LOLS model does not even improve as the sample size increases, thus indicating inconsistency. Indeed, it appears to get slightly worse in larger samples, which leads to the curious conclusion that, if one must use the LOLS model, collecting fewer observations might yield better results!

LOLS estimator is thus both empirically biased and inconsistent. Even if an infinite amount of data were available, the LOLS model would not provide meaningful estimates or even outperform an EPR model based on fewer observations.

An unbiased estimator is said to be "efficient" relative to the class of unbiased estimators if the variance of its estimates across samples is lower for all possible parameter values. Assessing statistical efficiency is useful, even when an estimator is biased. In fact, some biased estimators are preferred to their unbiased alternatives due to superior efficiency. In asymptotic samples, the EPR estimator is known to be efficient relative to all alternatives. The Monte Carlo experiments reported in Table 2 evaluate the efficiency of the EPR estimator relative to the LOLS estimator in realistic finite sample sizes.

For all 12 sample sizes and three coefficients, the standard deviations of the LOLS estimates are greater than those of the EPR estimates. Formal evaluation of relative efficiency is usually done in terms of a ratio of the two variances (squared standard deviations). This ratio indicates that in finite samples the EPR estimator is between 3.03 and 14.19 times as efficient as the LOLS estimator. As an example of how large this improvement is, note that Zellner's (1962) famous Seemingly Unrelated Regressions model improves efficiency by only about 0.20 times the alternative. Were we to evaluate mean square error (which combines both bias and variance in one expression), the improvement would be even more dramatic. All this means that the EPR model will be substantially better in practice than the LOLS model.

Thus, whether evaluated on the basis of unbiasedness, consistency, or efficiency, in large or small samples there is no reason to use the LOLS estimator. Although one can never know how general the results from Monte Carlo experiments are, this unambiguous evidence indicates that the coefficients do not have anything near the desirable properties. All provide quite imprecise estimates and biased test statistics, standard errors, and significance levels. These results also indicate that application of the EPR model will yield considerably better estimates and can lead to important changes in the substantive conclusions of empirical research.

The remaining Monte Carlo results focus on the EPR model and establish the value of its estimated standard errors and the sampling distribution of its estimators. The first row for each sample size in Table 2 reports the average estimated standard error from the EPR trials. Comparing this to the second row indicates that, for small sample sizes, the efficiency of the EPR estimates are consistently better than their estimated standard errors indicate. However, by sample sizes of

TABLE 3

Sampling Distribution of the EPR Coefficients

Sample Size	$\beta_1 = 2$	$\beta_2 = 0.4$	$\beta_3 = -3$
10	5.9924	3.7613	7.2224
20	0.2980	3.1980	3.7719
30	1.3836	2.3653	6.6104
40	1.6535	1.4663	5.8202
50	5.1830	1.8289	1.6881
100	1.9270	3.0000	1.0439
150	3.0370	8.7011	1.2592
200	8.6919	5.8943	3.4248
500	0.4929	5.2217	3.2574
750	3.1073	3.1621	3.2344
1,000	4.9751	4.0619	1.1826
2,000	2.7638	1.1700	1.1126

NOTE: Entries are chi-square tests ($d.f. = 5$) for normality of each set of 100 EPR coefficients for each sample size.

about 100, the two converge. Thus, for small sample sizes, standard errors and hypothesis tests will generally be conservative indicators.

Normality is guaranteed in infinite samples for any maximum likelihood estimator, but EPR estimates are not necessarily approximately normally distributed in finite samples. Normality is important to establish if test statistics and confidence intervals are to be used in the small samples usually available in political science research. Table 3 presents chi-square tests for the normality of the distribution of the EPR estimates across samples.[10] For each sample size, the 100 estimates are standardized (with mean equal to the population parameter) and the number that fall from $-\infty$ to -2, -2 to -1, -1 to 0, 0 to 1, 1 to 2, and 2 to ∞ are counted and compared to the number that are expected under an assumption of normality. A chi-square statistic ($d.f. = 5$) was calculated as $\sum_{i=1}^{6}(O_i - E_i)^2/E_i$, where O_i is the observed number of estimates and E_i is the expected number. If there were no error, chi-square of zero would indicate normality. Since random error is present, we use the expected value of each of these chi-square tests under the assumption of normality—which in this case is 5.0. In Table 3, 78 percent of the values are less than 5.0. Furthermore, the largest value in the table was only 8.7011; the unconditional probability of this one coefficient occurring with normality is 0.1216. These results indicate that the EPR coefficients can be con-

[10]The distribution of y (Poisson) and the distribution of the estimators, b, (normal) should not be confused. The central limit theorem guarantees normality of the coefficients for all maximum likelihood estimators regardless of the original distributional assumptions.

fidently used to test hypotheses and make probabilistic statements about the co-efficient estimates, even in very small samples.

One should always use caution when studying estimators with Monte Carlo methods, since the results are dependent upon a variety of experiment-specific choices. They are nevertheless quite useful in situations such as this. As Hendry (1984, p. 944) remarks, "A powerful advantage of experimental simulations is that much more *general* data generation processes or more complicated tech-niques can be investigated than can be tackled feasibly by analysis . . . [espe-cially for] investigating finite sample distributions."

7. An Empirical Application

In order to help with the interpretation of these models, an empirical ex-ample from political science research is presented. The number of members of the U.S. House of Representatives who switched political parties for each year, 1802–76, serves as the event count dependent variable to be analyzed (see King and Benjamin, 1985, for an extended justification of the theory, data, and speci-fications). Previous researchers demonstrated that party switchers can be de-scribed as having a troubling level of ideological or political dissonance that is

TABLE 4

Variable Descriptions

HSwitch	The number of members of the House of Representatives who switch parties in each year, 1802–76 (the dependent variable)
Inflation	Proportional change in the consumer price index
Military	Rate of change in percentage of the population who are mili-tary personnel on active duty
NonEleYear	1 if nonelection year, 0 otherwise
Professional	Proportional change in the percentage of the House who leave Congress for electoral reasons
PresElect	Absolute rate of change in the popular vote for the winning presidential candidate four years before
PublicLaw	Proportional change in the number of public laws promulgated by the U.S. Congress
$PublicLaw_{t-2}$	PublicLaw lagged two years
$HSwitch_{t-2}$	Number of members of the House who switched parties two years before
$HSwitch_{t-4}$	Number of members of the House who switched parties four years before
SSwitch	Number of members of the Senate who switched parties
Constant	A constant term equal to a vector of ones
$\ln(N\text{Mem})$	The natural logarithm of the number of members of the House

TABLE 5
House Political Party Switchers: An Empirical Application

Variable	EPR		LOLS							
			$c = 0.00001$		$c = 0.01$		$c = 1.0$		$c = 5.0$	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Inflation	−3.785	2.253	−16.960	8.875	−7.231	3.811	−1.450	0.889	−0.454	0.337
Military	0.401	0.259	2.257	1.307	0.972	0.562	0.200	0.131	0.066	0.050
NonEleYear	1.051	0.320	3.469	1.278	1.569	0.549	0.383	0.128	0.131	0.049
Professional	−1.986	0.667	−4.581	3.853	−2.465	1.655	−0.965	0.386	−0.428	0.147
PresElect	1.379	1.010	1.538	3.995	0.769	1.716	0.260	0.400	0.107	0.152
PublicLaw	0.206	0.365	0.320	2.531	0.099	1.087	−0.026	0.254	−0.034	0.096
PublicLaw$_{t-2}$	1.164	0.383	2.972	2.478	1.405	1.064	0.404	0.248	0.154	0.094
HSwitch$_{t-2}$	0.018	0.052	−0.263	0.396	−0.086	0.170	0.014	0.040	0.016	0.015
HSwitch$_{t-4}$	0.143	0.060	1.201	0.377	0.525	0.162	0.111	0.038	0.034	0.014
SSwitch	0.457	0.108	1.179	0.840	0.612	0.361	0.229	0.084	0.106	0.032
Constant	−20.230	6.045	−59.800	20.810	−25.450	8.937	−4.284	2.085	0.302	0.791
ln(N Mem)	3.490	1.097	9.452	3.824	4.062	1.642	0.818	0.383	0.246	0.145

resolved by switching parties. King and Benjamin therefore assumed that although some legislators always exist at or near the margin of switching parties, these people only make the jump at certain times. Based on a theory of political loyalty, the specific hypothesis was that large increases in environmental stress and pressure on the legislators resulted in relatively small increases in the frequency of party switching.

Table 4 provides a brief account of the variables used in this analysis. One problem with these data is that the number of members of the House varies over the time period analyzed. Thus, instead of the usual specification, it seems more reasonable to consider: $\theta_i/N\text{Mem}_i = \exp(x_i\beta)$. This is appropriate and can be analyzed by recognizing that it is the same specification as (i.e., algebraically equivalent to) $\theta_i = \exp[x_i\beta + \ln(N\text{Mem}_i)]$. This implies that $\ln(N\text{Mem}_i)$ should be included as an explanatory variable with its coefficient constrained to one. However, freeing it up, by just including it as another explanatory variable, can do no real harm (see Maddala, 1983, p. 52). This procedure also has intuitive appeal, since we are actually controlling for the size of the House.

Table 5 reports the results of an EPR estimation and four LOLS estimations, with the small constant c taking on the values 0.00001, 0.01, 1.0, and 5.0. I concentrate first on interpreting several of the EPR coefficients and then move to a comparison of the EPR and LOLS estimations.[11]

Inflation in this early period was a positive economic event; America was a debtor nation and deflation was feared (indeed, not until the late nineteenth and especially twentieth centuries was inflation viewed as something to be avoided). Thus, as expected, the coefficient from the EPR estimation is negative (-3.785), indicating that more inflation led to fewer party switches. To interpret these results more precisely, recall that, as set up in section 3, the effect of x on y is $\beta\theta$. As a result, when the expected number of switchers is 1.347 (the empirical mean of the dependent variable), the effect of inflation is $-3.785 \times 1.347 = -5.098$. Inflation is measured here as a proportionate (not percentage) change; as a result, when it takes on a value of 1, inflation has increased by 100 percent. In this unlikely situation, the number of switchers would drop by about five, controlling for the other variables in the equation. When the inflation rate increases by 0.20 (20 percentage points), the effect is $-5.098 \times 0.20 = -1.02$, indicating that about one fewer representative will switch parties. Suppose now that the expected number of switches was, say, five due to an already quite stressful year. In this case, a 0.20 increase in inflation would lead on average to about $-3.785 \times 5 \times 0.20 = -3.785$, that is nearly four fewer switchers. This larger effect is consistent with the theoretical development of the EPR model. The same change in an explanatory variable is likely to have a larger effect when the expected value of the dependent variable is already further from zero, its lower bound.

[11] Note that t-tests and probability values can be calculated in the same manner as with correctly estimated OLS coefficients: dividing the estimate by the standard error for each coefficient provides a t-test for the hypothesis that the true parameter is zero.

Consider also NonEleYear, a dummy variable indicating whether it is an election (0) or nonelection (1) year. If the expected number of switchers is 1, then about 1.051 more representatives switch parties in nonelection than election years, holding the other variables in the equation constant. (This is plausible, since switching has tremendous potential electoral implications; planning for the next election becomes considerably more difficult and time consuming.) For years in which the expected number of switches is higher, say, 10, the model indicates that about $10 \times 1.051 = 10.51$ more representatives will switch in nonelection years than in election years.

Observe now the comparison between the EPR and the first two sets of LOLS coefficients (with $c = 0.00001$ and $c = 0.01$).[12] Note that, in all but one case for $c = 0.00001$ and two cases for $c = 0.01$, the LOLS coefficients are larger in absolute value than the EPR coefficients. This is consistent with the Monte Carlo results in the previous section and with the analytical results in section 5. Also consistent with section 5's results is that the size of the LOLS coefficients drops rather quickly—even below the EPR estimates—as c increases. Thus, whereas the Inflation coefficient is -3.785 under the EPR model, the LOLS estimates range from -16.960 to -0.454; the arbitrary choice of a value for c can therefore have enormous consequences for one's substantive conclusions.[13]

Indeed, although the analytical results are reflected in the general patterns found in this table, other types of problems are also present. For example, the positive EPR coefficient on $HSwitch_{t-2}$ indicates that switching two years ago leads to more switching today (although the coefficient is small and conventionally nonsignificant). As hypothesized, this reflects an emulation or diffusion effect. While two of the LOLS estimates for this variable are also positive, two others have changed sign, indicating precisely the opposite substantive conclusion! Also problematic is that as c changes, the standard errors change as dramatically as the coefficients.

8. Summary and Conclusions

There are three major conclusions of this study of event count data.
1. The underlying mathematical process generating event count data is

---

[12]Although a direct comparison of these LOLS and EPR coefficients with OLS coefficients estimated from the same data is inappropriate, the fitted values from the OLS model can yield useful information. In this data set, 12 of 75 fitted values were less than zero; this nonsensical result provides a strong indication that there is a problem with the linear model in this case.

[13]On the very odd chance that one were to guess the "best" value for c, I found the value of c in this example for which the sum of squared differences between the LOLS and EPR coefficients was smallest. For every empirical example to which I have applied this procedure, different "best" values of c were found. In the present case, $c = 0.025$ is "best." However, even in this case one of the signs is different, and some of the coefficients are still of quite different sizes. The sum of squared differences was 16.26, quite a large amount of error. Furthermore, that the variance of each LOLS coefficient was still considerably larger (from 1.5 to 7.7 times larger) than its respective EPR coefficient variance indicates a substantial loss of efficiency in the LOLS model.

driven by the Poisson distribution. This process applies to nearly all time series and cross-sectional event counts common in many areas of political science research. The evidence for event count data being described in this way is considerably stronger than the evidence that most dependent variables used in political science regression analyses follow a normal distribution.

2. When commonly used methodologies such as ordinary least squares or logged OLS are applied to data that are distributed as Poisson, very serious problems can and usually will result. The conditional expectation function has no finite or meaningful evaluation. The estimators are inefficient, insufficient, inconsistent, and biased, even in infinite samples. Experiments, using the commonly used $c = 0.01$ constant, demonstrate that application of these models yields coefficients biased upward in absolute value and quite imprecise as well. In a political application, the analytical and Monte Carlo results were borne out, indicating very large biases and an extreme sensitivity to the choice of a small constant added to y so that logs could be taken.

3. The exponential Poisson regression model provides an unbiased and consistent alternative estimator which, in large samples, is more efficient than any other. Monte Carlo experiments indicated that these and other very desirable statistical properties seem to exist even in samples as small as 10.[14]

The differences between the widely used conventional methods and the exponential Poisson regression (EPR) model are not minor technical discrepancies. In large samples, the former literally make very little sense, whereas the latter is very plausible. In the small samples commonly found in political science, dramatically improved estimates can result from the application of the EPR model. The formal statistical problems may appear esoteric to some applied researchers, but in this case they have real substantive consequences. Empirical estimates can be many times as large as they should be; in applications, even the signs of the coefficients can be incorrect in the OLS and LOLS models, while correct for the EPR model. Adopting the EPR model would not only make one technically correct, but it also would be more likely to produce meaningful data analyses, would help to find previously undiscoverable relationships, would probably make some dramatic changes in existing substantive literature, and would encourage work in areas where many relationships appeared "not significant" or too imprecise or sensitive to warrant further analysis or publication.

Manuscript submitted 26 June 1986
Final manuscript received 9 June 1987

[14] When more than one parallel series of event counts are available, a seemingly unrelated Poisson regression model can be analyzed (see King, 1987b).

APPENDIX A
Proof That Event Counts Are Poisson Processes [16]

Begin by writing the probability of an addition, and of no addition, respectively, to the total count during the interval from t to $t + \Delta t$ as:

$$Pr(y_{t+\Delta t} = n + 1 \mid y_t = n) = \theta \Delta t + o(\Delta t) \tag{A.1}$$
$$Pr(y_{t+\Delta t} = n \mid y_t = n) = 1 - [\theta \Delta t + o(\Delta t)]$$

where $o(\Delta t)$ is the probability that more than one event occurs during Δt and which, when divided by Δt, tends to zero as Δt gets smaller. We can then write the unconditional probability $Pr(y_{t+\Delta t} = n + 1)$ as the sum of two mutually exclusive situations: (1) n events have occurred by time t and one additional event occurs over the next Δt interval, and (2) $n + 1$ events have occurred at time t and no new events occur from t to $t + \Delta t$. Dropping the $o(\Delta t)$ terms from A.1 (i.e., assuming that two events cannot occur at precisely the same instant), the probability of each of these situations can be written as the product of two marginal probabilities due to the Markov assumption that the probability of event occurrence at the two times are independent:

$$Pr(y_{t+\Delta t} = n + 1) = Pr(y_t = n)\theta \Delta t + Pr(y_t = n + 1)(1 - \theta \Delta t) \tag{A.2}$$

From A.2, observe how $Pr(y_t = n + 1)$ changes with respect to time as Δt gets smaller and smaller:

$$\frac{\partial Pr(y_t = n + 1)}{\partial t} = \lim_{\Delta t \to 0} \frac{Pr(y_{t-\Delta t} = n + 1) - Pr(y_t = n + 1)}{\Delta t}$$
$$= \begin{cases} \theta[Pr(y_t = n) - Pr(y_t = n + 1)] & \text{for} \quad n + 1 > 0 \\ -\theta Pr(y_t = 0) & \text{for} \quad n + 1 = 0 \end{cases} \tag{A.3}$$

If the event count is set to zero at the start of the period, then $Pr(y_0 = 0) = 1$. As such, the probability distribution of this underlying process can begin to be built. First, solve the last part of A.3 as $Pr(y_t = 0) = e^{-\theta t}$, since exponentiation is the only function that is equal to its derivative. Then substituting into the other part of A.3 yields $Pr(y_t = 1) = \theta t e^{-\theta t}$. Finally, successively substituting and solving yields the general formula for the Poisson distribution with parameter θt (Figure 1 was drawn from this equation):

$$Pr(y_t = n) = \frac{e^{-\theta t}(\theta t)^n}{n!} \quad \text{for} \quad n = 0, 1, 2, \ldots \tag{A.4}$$

If all observation periods are the same length, t is set to 1.0 in this equation.

APPENDIX B
Computer Algorithms for the EP Regression Model

The most widely available program set up to estimate the exponential Poisson regression model is GLIM (Generalized Linear Interactive Modeling; The Numerical Algorithms Group, 7 Banbury Road, Oxford OX2 6NN). Some versions of the SAS computer program (SAS Institute, Box 8000, Cary, NC 27511) have also incorporated the GLIM commands. With the following code, GLIM can perform an EPR estimation on 100 observations with Y, as the dependent variable, and $X1$, $X2$, and $X3$, as the independent variables. {Comments below are indicated by braces and are not to be entered in the program.}

```
GLIM              {Operating system command to execute GLIM}
$UNITS 100        {Specify number of observations}
```

[16] This proof relies on insights into the continuous time, discrete space Markov process outlined by Feller (1968, ch. 17).

$DATA Y $X1$ $X2$ $X3$	{List all variable names for free format input data}
$DINPUT 22	{Read input data from file 22}
$ERROR P	{Specify disturbances as Poisson distributed}
$LINK L	{Specify a "log link." That is, $\ln(\theta) = X\beta$ or equivalently $\theta = \exp(X\beta)$.}
$YVAR Y	{Identify the dependent variable}
$FIT $X1 + X2 + X3$	{Fits EPR model. A constant term is included by default.}
$DISPLAY M E	{Displays the model and estimates}

The authors of SST (Dubin/Rivers Research, 1510 Ontario Avenue, Pasadena, CA 91103) have recently incorporated EPR into new versions of their program. SST is an excellent combination of versatility and very simple command syntax for microcomputers. For EPR, do:

—range obs[1 − 100]
—read file[file.dat] to[y X1 X2 X3]
—set one = 1
—Poisson dep[y] ind[one X1 X2 X3]

Other computer programs can estimate any model for which a log-likelihood function can be written. The log-likelihood function for the EPR model is given in equation 3.3 (note 4). In some computer programs, all that needs to be done is to enter this equation. In others, the first and second derivatives of this function must be entered. In still others, the derivatives are optional but will speed up estimation considerably in most problems. Fortunately, both derivatives take quite simple forms:

$$\frac{\partial \ln(L_i)}{\partial \beta} = -x_i \exp(x_i' \beta)$$

$$\frac{\partial^2 \ln(L_i)}{\partial \beta \partial \beta'} = -x_i^2 \exp(x_i' \beta)$$

The GAUSS computer program (ApTech Systems, P.O. Box 6487, Kent, WA 98064) implements EPR in this way. GAUSS may be too technical for the occasional user, but it is the most flexible program available. The code necessary to run an EPR model is available from the author of this paper.

LIMDEP (available from its author, Professor William H. Greene, Graduate School of Business Administration, New York University, 100 Trinity Place, New York, NY 10006) is also relatively technical but is still a good general choice for a variety of problems on mainframes and microcomputers. Section 40.9 of the LIMDEP manual demonstrates how to estimate an exponential Poisson regression model.

REFERENCES

Allison, Paul D. 1984. *Event history analysis: Regression for longitudinal event data.* Beverly Hills: Sage.

Amburgey, Terry L., and Glenn R. Carroll. 1984. Time-series models for event counts. *Social Science Research,* 13:38–54.

Azar, Edward E., and Thomas J. Sloan. 1975. *Dimensions of interaction.* Occasional Paper No. 8, Pittsburgh: International Studies Association.

Berndt, E. K., B. H. Hall, R. E. Hall, and J. A. Hausman. 1974. Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement,* 3:653–65.

Campbell, James E. 1985. Explaining presidential losses in midterm congressional elections. *Journal of Politics,* 47:1140–57.

Chaing, Alpha. 1974. *Fundamental methods of mathematical economics.* New York: McGraw-Hill.

Cook, Timothy. 1986. House members as newsmakers: The effects of televising Congress. *Legis-lative Studies Quarterly,* 2:203–26.

Di Franceisco, Wayne, and Zvi Gitelman. 1984. Soviet political culture and "covert participation" in policy implementation. *American Political Science Review,* 78:603–21.

El-Sayyad, G. M. 1973. Bayesian and classical analysis of Poisson regression. *Journal of the Royal Statistical Association,* series B35.3 (Methodological):445–51.

Feller, William. 1968. *An introduction to probability theory and its applications.* Vol. 1, 3rd ed. New York: Wiley.

Gourieroux, C., A. Monfort, and A. Trognon. 1984. Pseudo maximum likelihood methods: Theory. *Econometrica,* 52:681–700.

Haight, F. 1967. *Handbook of the Poisson distribution.* New York: Wiley.

Hanson, Russell. 1984. Medicaid and the politics of redistribution. *American Journal of Political Science,* 28:313–39.

Hendry, David F. 1984. Monte Carlo experimentation in econometrics. Zvi Griliches and Michael D. Intriligator, eds., *Handbook of Econometrics.* Vol. 2. New York: North-Holland.

Johnson, Norman L., and Samuel Kotz, 1969. *Discrete distributions.* New York: Wiley.

Johnson, Thomas H., Robert O. Slater, and Pat McGowan. 1984. Explaining African military coups d'etat, 1960–1982. *American Political Science Review,* 78:622–40.

Jorgenson, Dale W. 1961. Multiple regression analysis of a Poisson process. *Journal of the American Statistical Association,* 56:235–45.

King, Gary. 1986. How not to lie with statistics: Avoiding common mistakes in quantitative political science. *American Journal of Political Science,* 30:666–87.

———. 1987a. Presidential appointments to the Supreme Court: Adding systematic explanation to probabilistic description. *American Politics Quarterly,* 15:373–86.

———. 1987b. A seemingly unrelated Poisson regression model. Presented at the annual meeting of the Political Science Methodology Group, Harvard University.

———. 1987c. Variance specification in event count models: From restrictive assumptions to a gen-eralized estimator. Presented at the annual meeting of the Political Science Methodology Group, Duke University.

———. Forthcoming. *Creative statistical modeling: The likelihood theory of inference.*

King, Gary, and Gerald Benjamin. 1985. The stability of party identification among U.S. senators and representatives: 1789–1984. Presented at the annual meeting of the American Political Science Association, New Orleans.

Kritzer, Herbert M. 1978. Analyzing contingency tables by weighted least squares: An alternative to the Goodman approach. *Political Methodology,* 5:277–326.

Maddala, G. S. 1977. *Econometrics.* New York: McGraw-Hill.

———. 1983. *Limited dependent and qualitative variables in econometrics.* Cambridge: Cambridge University Press.

McAdams, John C., and John R. Johannes. 1985. Constituency attentiveness in the House: 1977–82. *Journal of Politics,* 47:1108–39.

McCullagh, P., and J. A. Nelder. 1983. *Generalized linear models.* London: Chapman and Hall.

Nelder, J. A., and R. W. M. Wedderburn. 1972. Generalized linear models. *Journal of the Royal Statistical Society,* series A, 135:370–84.

Plackett, R. L. 1981. *The analysis of categorical data.* 2nd ed. New York: Macmillan.

Poisson, Simeon Denis. 1837. *Recherches sur la probabilite' des jugements en matiere criminelle et en matiere civile, precedees des regles generales du calcul des probabilites.* Bachelier, Impri-meur-Libraire pour les Mathematiques, la Physique, etc. Quai des Augustins, No. 55.

Rohde, David W., and Dennis M. Simon. 1985. Presidential vetoes and congressional response: A study of institutional conflict. *American Journal of Political Science,* 29:397–427.

Robertson, John D. 1984. Toward a political-economic accounting of the endurance of cabinet ad-

ministrations: An empirical assessment of eight European democracies. *American Journal of Political Science,* 28:693–709.

Sigelman, Lee, and Dixie Mercer McNeil. 1980. White House decision making under stress: A case analysis. *American Journal of Political Science,* 24:652–73.

Tuma, Nancy Brandon, and Michael T. Hannan. 1979. Dynamic analysis of event histories. *American Journal of Sociology,* 84:820–54.

Zellner, Arnold. 1962. An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias. *Journal of the American Statistical Association,* 57:348–68.