Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator*

Gary King, *Harvard University*

This paper discusses the problem of variance specification in models for event count data. Event counts are dependent variables that can take on only nonnegative integer values, such as the number of wars or coups d'etat in a year. I discuss several generalizations of the Poisson regression model, presented in King (1988), to allow for substantively interesting stochastic processes that do not fit into the Poisson framework. Individual models that cope with, and help analyze, heterogeneity, contagion, and negative contagion are each shown to lead to specific statistical models for event count data. In addition, I derive a new generalized event count (GEC) model that enables researchers to extract significant amounts of new information from existing data by estimating features of these unobserved substantive processes. Applications of this model to congressional challenges of presidential vetoes and superpower conflict demonstrate the dramatic advantages of this approach.

1. Introduction

Event counts are dependent variables that take on nonnegative integer values for each of n observations. These values represent the number of times an event occurs in a fixed domain. The domain for each observation may be a cross-section, a time interval, or even the cell of a contingency table. The number of visible uses of military force initiated by the United States in each six-month interval (Stoll, 1984), the number of presidential vetoes per year (Rohde and Simon, 1985), the frequency of formal and informal military alliances (Russett, 1971; McGowan and Rood, 1975), and the annual number of presidential appointments to the Supreme Court (King, 1987) are examples of time series counts. Examples of cross-sectional event count studies include the number of coups d'etat in each black African state (Johnson, Slater, and McGowan, 1984) and the number of political activities engaged in and reported by Soviet emigres (Di Franceisco and Gitelman, 1984).[1]

[1]Event counts may also vary across both time and space, but these require special pooled time series–cross-sectional models not considered here. Different models of this sort have been

Least squares analyses of event counts are very inefficient, have inconsistent standard errors, and may produce negative predictions for the number of events; least squares estimates with a logged dependent variable suffer from these problems and are biased and inconsistent as well (King, 1988). The Poisson regression model is now used in many disciplines in lieu of these models, but it makes two key assumptions about the way unobserved processes generate event counts that are implausible in many applications. Providing estimates of these unobserved processes, instead of assuming them, can lead to important insights about empirical data. If these assumptions do not hold, but the Poisson model is applied anyway, parameter estimates will be inefficient and standard errors inconsistent, a situation analogous to heteroscedasticity in least squares models. In fact, this problem in event count models is generally more serious, since the inconsistency in the standard errors is quite dramatic and since the problems can sometimes lead to inconsistency in the parameter estimates as well.

In this paper I derive a new generalized event count model. This model enables researchers to estimate what was previously assumed and to study much more interesting data generation processes. Perhaps most striking is that the model extracts significant additional information from existing data and substantially improves estimation without *any* cost in additional assumptions. Section 2 contains a review of the basic Poisson model with an emphasis on its variance assumptions. In section 3 the processes giving rise to overdispersion (the variance being greater than the mean) are analyzed. In section 4 processes leading to underdispersion are analyzed, and a new maximum likelihood estimator is proposed. However, limited prior information, and, hence, an unknown range of dispersion, is the usual situation with political event count analyses. Thus, a new generalized event count estimator is proposed in section 5. This estimator is consistent in the presence of unknown forms and levels of under-, over-, or Poisson dispersion. Since this new and more general estimator is almost always more appropriate in actual research situations, sections 3 and 4 should be considered important primarily as a lead up to section 5. Section 6 provides two examples that demonstrate the dramatic advantages in political science research that can result from the application of this new model compared to the more commonly used Poisson model. Section 7 concludes.

---

applied to the number of patents awarded per firm each year (Hausman, Hall, and Griliches, 1984) and the number of presidential vetoes in each policy area per year (King, 1989b). See also Bishop, Fienberg, and Holland (1975), McCullagh and Nelder (1983), and King (1989c, ch. 6) on log-linear models for contingency tables. These models can be considered special cases of the event count models discussed below.

2. The Poisson Regression Model

Let Y_i $(i = 1, \ldots, n)$ be the random dependent variable event count, so that only the values $0, 1, \ldots$, occur with nonzero probability. We observe the realization y_i only at the end of each observation period i. To derive a specific probability distribution, one nevertheless needs to make specific assumptions about the *unobserved process* within each observation period generating the *observed count* at the end of the period. For example, suppose we make the following assumptions about the process during observation period i:

ASSUMPTION 1: More than one event cannot occur at the same instant.

ASSUMPTION 2: The probability of an event occurring at any instant is *constant* within period i and *independent* of all previous events during that observation period.

ASSUMPTION 3: Zero events have occurred at the start of the period.

From these first principles, one can derive a form of the Poisson probability distribution for the random variable Y_i (see King, 1988, Appendix 1):

$$f_{pt}(y_i \mid \lambda_i, t_i) = \begin{cases} \dfrac{e^{-\lambda t_i}(\lambda t_i)^{y_i}}{y_t!} & \text{for } t_i > 0, \lambda_i > 0 \text{ and } y_i = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The more usual form of the Poisson distribution results from adding the additional first principle:

ASSUMPTION 4: The length t_i of each observation period i is identical (and equal to, say, 1.0).

Thus, we have:

$$f_p(y_i \mid \lambda_i) = \begin{cases} \dfrac{e^{-\lambda_i}(\lambda_i)^{y_i}}{y_i!} & \text{for } \lambda_i > 0 \text{ and } y_i = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

An event count regression model is specified by letting the expected count, $E(Y_i) \equiv \lambda_i$ in one of these equations vary over observations according to a specific function of a vector of explanatory variables. This may be written in general form as $\lambda_i = \lambda_i(x_i, \beta)$, where x_i is a vector of k exogenous variables and β is a $k \times 1$ parameter vector. In this paper I specialize this functional form to $E(Y_i) \equiv \lambda_i = \exp(x_i\beta)$ for the same theoretical reasons developed and justified in King (1988). I shall use this more specific form for each of the models discussed below, but, at any time before differentiating, another form may be substituted for λ_i.

To estimate β, the effect of the explanatory variables on the dependent variable, the method of maximum likelihood is used. By assuming the absence of autocorrelation, the likelihood function may be written as

$$L(\beta \mid y) = \prod_{i=1}^{n} f_p(y_i \mid \lambda_i) = \prod_{i=1}^{n} \frac{e^{-\lambda_i}(\lambda_i)^{y_i}}{y_i!} \tag{3}$$

with $\lambda_i = \exp(x_i\beta)$. The basic idea of maximum likelihood is that the data are taken as given, and we choose the value of β that maximizes this function and hence the likelihood of having generated the data (see King, 1989c, for an introduction). The log-likelihood function, reduced to sufficient statistics, is then

$$\ln L(\beta \mid y) = \sum_{i=1}^{n} \{-\exp(x_i\beta) + y_i(x_t\beta)\} \tag{4}$$

Standard numerical maximization methods can easily be applied to this globally concave function by using one of many available computer programs (see note 11 below and King, 1988, Appendix 2).

The focus of this paper is on the four assumptions of the stochastic process generating these counts. Assumption 1 is a relatively technical requirement without many real consequences. Assumption 3 is more of a notational convenience, and the presence of Assumption 4 enables one to choose among the two distributions in equations 1 and 2. However, for political science research, Assumption 2 is very important. For example, suppose one is analyzing the number of coups d'etat. Observation i is a single black African state over some period (say, 20 years). The data only contain information on the total number of coups that have occurred in country i; they do not indicate in what way, during the 20-year period, the coups occurred.[2] Assumption 2 has two implications for this example. In one, the probability of a coup is assumed the same for every day during those 20 years. This *homogeneity* assumption seems quite implausible. Economic conditions, civil unrest, international conflict, and other factors are likely to vary over time, resulting in a heterogeneous probability of event occurrence. Second, the occurrence of a coup at one point in time is assumed *independent* of all coups that might have occurred before. This part of the assumption is also implausible in this example because one coup is likely to decrease the probability of a coup in the very near future, but a successful coup may also increase the probability of a coup farther along.

---

[2]If this more detailed information existed, one could use event history methods and model these processes in the systematic, rather than stochastic, component of a statistical model (see Allison, 1984). Unfortunately, this detailed information is often either difficult or impossible to obtain for all one's dependent and independent variables.

These assumptions about the unobserved process generating the observed counts have consequences for the variance in the event count regression. The variance of Y_i under the Poisson distribution in equation 2 is equal to its expected value:

$$V(Y_i) = E(Y_i) \equiv \lambda_i \tag{5}$$

However, this result at the level of the counts relies on microlevel assumptions about the process generating the counts—independence and homogeneity. If these assumptions do not apply, then the Poisson distribution does not result, and the variance is not equal to the mean. In this case the log-likelihood in equation 4 will yield consistent estimates, but they will be inefficient, and the standard errors inconsistent.

More generally, let

$$V(Y_i) = \lambda_i \sigma^2 \tag{6}$$

for $\lambda_i > 0$ and $\sigma^2 > 0$; σ^2 is called the *dispersion parameter*. When individual events are independent with homogeneous rates of occurrence, the result is $\sigma^2 = 1$ and $V(Y_i) = \lambda_i$, the case of Poisson dispersion. Alternative assumptions lead to other values for σ^2. For example, if $\sigma^2 > 1$, the data are said to be *overdispersed* and, if $0 < \sigma^2 < 1$, *underdispersed* (the next two sections describe the processes leading to these two situations). In general, $\sigma^2 = \sigma^2(x_i, \beta, \theta)$, where θ is now the ancillary parameter. Numerous functional forms for $V(Y_i)$ are thus possible (see Cameron and Trivedi, 1986, p. 33), but I use the more specialized form where σ^2 in equation 6 is a scalar parameter. In any of the models below, this form can easily be changed by substituting the alternate form in for σ^2 before differentiation. Furthermore, "even relatively substantial errors in the assumed functional form of [the variance] generally have only a small effect on the conclusions" (McCullagh and Nelder, 1983, p. 132).

3. Modeling Event Counts with Overdispersion

Heterogeneity and contagion, two very different and substantively interesting unobserved processes, can produce identical models for overdispersion ($\sigma^2 > 1$ in equation 6). Suppose the number of events recorded at each observation is the aggregation of a large number of individual units, each at risk λ_i for one event occurrence (e.g., persons in county i each at risk of committing suicide; bills sent to the president in year i each at risk of being vetoed). The aggregation process produces the realized number of events that occurred during the period, y_i (e.g., suicides or vetoes per year) and a random variable, Y_i, described by a complete probability distribution; this distribution will generally have an expected number of events, $E(Y_i) \equiv \lambda_i = \exp(x_i\beta)$ and a variance, $V(Y_i) = \lambda_i\sigma^2$. As noted above the

distribution will be Poisson if all micro events have equal probabilities of occurrence and are independent, and thus $\sigma^2 = 1$. For example, in any one day, the probability of any airplane in the domestic United States being hijacked is essentially identical. In addition, since information of one hijacking probably takes at least a day to be transmitted, we can expect that one hijacking incident would not influence the probability that other planes would be hijacked in the same day. Thus, the number of hijackings plausibly follows a Poisson distribution (see Holden, 1987), and thus σ^2 is probably 1.0.

However, if individual units within observation i are heterogeneous, λ_i will vary more across individual units (within observation i), and overdispersion will result.[3] For example, if we included all airplanes worldwide, some nations certainly have higher probabilities of hijacking occurrences than others.

In order to model *heterogeneous* processes, resulting in overdispersed event counts, I drop the assumption that λ_i is constant within observation i. Instead, I assume that it is a random variable. In order to build a stochastic model for heterogeneous processes, we must make some assumption about how λ_i (e.g., the rate of hijacking) is distributed across the micro units (airplanes) within each observation period (day). The usual assumption is that λ_i follows a gamma distribution (Johnson and Kotz, 1970, ch. 17). Under the gamma distribution $[f_\gamma(\lambda_i \mid \phi_i, \sigma^2)]$, the random variable λ_i takes on only nonnegative real numbers and is assumed to have mean $E(\lambda_i) \equiv \phi_i$ and variance $V(\lambda_i) \equiv \sigma^2$. The form of this distribution is quite flexible and is thus not overly restrictive, but one must recognize that this is a particular assumption about the nature of the unobserved heterogeneity. Other assumptions are possible, but do not reduce to closed form, and thus require more complicated estimation procedures (see Schoenberg, 1985).

Greenwood and Yule (1920) first derived the new distribution, called the negative binomial, by adding this additional first principle (λ_i following a gamma distribution) to the initial four with the following procedure. First, derive the joint distribution (f_j) of Y_i and λ_i (both now random variables) using the basic rule for conditional probability $[\Pr(AB) = \Pr(A \mid B)\Pr(B)]$:

$$f_j(y_i, \lambda_i \mid \phi_i, \sigma^2) = f_p(y_i \mid \lambda_i)f_\gamma(\lambda_i \mid \phi_i, \sigma^2) \tag{7}$$

Then, one derives the negative binomial distribution (f_{nb}) by collapsing this joint distribution over λ_i:

---

[3]Apparent overdispersion can be caused by measurement error in the explanatory variables (Prentice, 1986, p. 324) or the omission of relevant explanatory variables (uncorrelated with the ones included).

$$f_{nb}(y_i \mid \phi_i, \sigma^2) = \int_{-\infty}^{\infty} f_j(y_i, \lambda \mid \phi_i, \sigma^2) d\lambda \tag{8}$$

In the negative binomial distribution, the left-hand side of equation 8, the parameter ϕ_i plays the same role of the mean rate of event occurrence as λ_i does in the Poisson distribution. Thus, to maintain comparability, I reparameterize by substituting λ_i for each occurrence of ϕ_i and write out the entire distribution:

$$f_{nb}(y_i \mid \lambda_i, \sigma^2) = \frac{\Gamma\left(\dfrac{\lambda_i}{\sigma^2 - 1} + y_i\right)}{y_i! \Gamma\left(\dfrac{\lambda_i}{\sigma^2 - 1}\right)} \left(\frac{\sigma^2 - 1}{\sigma^2}\right)^{y_i} (\sigma^2)^{\frac{-\lambda_1}{\sigma^2 - 1}} \tag{9}$$

where $\lambda_i > 0$ and $\sigma^2 > 1$ and $\Gamma(\cdot)$ is the gamma function.

The result in equation 9 is a probability distribution with an additional parameter. The expected number of events can be modeled as before, $E(Y_i) \equiv \lambda_i = \exp(x_i \beta)$. However, the variance is now greater than the mean since

$$V(Y_i) = \lambda_i \sigma^2 = \exp(x_i \beta) \sigma^2 \tag{10}$$

and $\sigma^2 > 1$. As σ^2 approaches one, this distribution approximates the Poisson distribution. Larger values of σ^2 produce a distribution with larger and larger amounts of overdispersion in the counts, resulting from more heterogeneity within each observation.

Contagion is another process that also leads to overdispersion. Contagion occurs when the expected number of events at one time is dependent on the realized number of events at some previous time. For example, a single act of political terrorism is likely to stimulate future terrorist acts. For another example, one conflictual act between the United States and the Soviet Union is likely to·yield a flurry of tit-for-tat behavior. Since with event count data we only observe the total number of events at the end of the period, contagion, like heterogeneity, is an unobserved, within-observation process.[4]

Two distributions designed to model this sort of contagion are the contagious Polya-Eggenberger distribution and Neyman's contagious distributions. A remarkable result due to Thompson (1954) is that a limiting form of both distributions is the same negative binomial that was derived above for a heterogeneous event count process. For research problems where both heterogeneity and contagion are plausible, the different underlying pro-

---

[4]One can model cross-observation dependence by including a lagged value of y_i as an independent variable or through more complicated procedures.

cesses are not distinguishable with aggregate event count data because they both lead to the same probability distribution for the counts (see Neyman, 1965, p. 5). One can still use this distribution to derive fully efficient and consistent estimates, but this analysis will only be suggestive of the underlying process. In other situations, where a researcher can rule out one of these processes, this result is quite convenient for analyzing event counts, since the same model can be used.

A negative binomial maximum likelihood solution yields consistent and fully efficient parameter estimates in the case of overdispersion due to contagion or heterogeneity. The log-likelihood is then as follows:

$$
\ln L(\beta, \sigma^2 \mid y) = \sum_{i=1}^{n} \left\{ \ln \Gamma \left(\frac{\lambda_i}{\sigma^2 - 1} + y_i \right) - \ln \Gamma \left(\frac{\lambda_i}{\sigma^2 - 1} \right) \right.
$$
$$
\left. + y_i \ln(\sigma^2 - 1) - \ln(\sigma^2) \left(y_i + \frac{\lambda_i}{\sigma^2 - 1} \right) \right\}
$$
(11)

where $\lambda_i = \exp(x_i\beta)$, for $y_i = 0, 1, 2, \ldots$ and $\sigma^2 > 1$.

Whereas in the Poisson regression model one maximizes the log-likelihood with respect to β, this log-likelihood is maximized with respect to both β and σ^2. The information from the maximum likelihood estimate, $\hat{\beta}$, is interpreted as with the Poisson model. In this case the maximum likelihood estimate, $\hat{\sigma}^2$, provides information about how overdispersed (heterogeneous or contagious) the data are. Both $\hat{\beta}$ *and* $\hat{\sigma}^2$ can be useful evidence when attempting to extract information from event count data.

4. Modeling Event Counts with Underdispersion

Models for underdispersion have generally been neglected in the event count literature. Some work exists on truncated event count models that produce underdispersion, but these apply only to more special cases where the untruncated parent distribution is Poisson or overdispersed (Mullahy, 1986; King, 1989a). Statisticians sometimes argue that underdispersion is unusual in their data. I find this tendency in political science data as well, although in my experience examples of underdispersion can be found. Indeed, one can readily conceive of many interesting political science examples that could generate underdispersed data. *Negative contagion,* where the occurrence of one event makes future events somewhat less likely, is generally the microlevel cause of underdispersion. For example, when one politician becomes a candidate for his or her party's presidential nomination, other potential candidates may be less likely to enter the race. Negative contagion also characterizes the situation where events are very unlikely to occur immediately following other events of the same type. For example, when some congressional Democrats decided to attempt to block the elevation of William Rehnquist from associate to chief justice of the U.S. Su-

preme Court, they decided not to attempt to block the confirmation of An-
tonin Scalia as associate justice, even though they opposed him as well. The
Democrats reasoned that success in opposing the second nomination would
be much less likely to be successful, given the very short interval of time be-
tween the two.

It is well known that the process of negative contagion can give rise to
a binomial probability distribution and underdispersion in the event counts
(e.g., Patil and Boswell, 1975). However, in the present context, this ap-
proach has a serious limitation. To illustrate this problem and provide a so-
lution, I first reparameterize the binomial distribution along the lines of the
Poisson and negative binomial distributions:

$$
f_b(y_i \mid \lambda_i, \sigma^2) = \frac{\Gamma\left(\dfrac{-\lambda_i}{\sigma^2 - 1} + 1\right)}{y_i!\,\Gamma\left(\dfrac{-\lambda_i}{\sigma^2 - 1} - y_i + 1\right)}(1 - \sigma^2)^{y_i}(\sigma^2)^{\frac{-\lambda_i}{\sigma^2 - 1} - y_i}
\tag{12}
$$

and where $E(Y_i) \equiv \lambda_i = \exp(x_i\beta) > 0$, $V(Y_i) = \lambda_i\sigma^2$, and $0 < \sigma^2 < 1$. This expression directly addresses the situation of underdispersion
but has one disqualifying feature. If $(-\lambda_i)/(\sigma^2 - 1)$ is not an integer,
$\Sigma_{m=0}^{\infty} f_b(m \mid \lambda_i, \sigma^2) \neq 1$; thus, it is not a proper probability distribution.

This problem can be solved by defining a new probability distribution,
which I call the "continuous parameter binomial." A version of this distri-
bution appears in Johnson and Kotz (1969, p. 40). The continuous param-
eter binomial may be defined as follows;

$$
f_{cpb}(y_i \mid \lambda_i, \sigma^2) = \frac{f_b(y_i \mid \lambda_i, \sigma^2)}{D_i}
\tag{13}
$$

where

$$
D_i = \sum_{m=0}^{\left[\frac{-\lambda_i}{\sigma^2 - 1}\right] + 1} f_b(m \mid \lambda_i, \sigma^2)
\tag{14}
$$

This distribution is defined on the integers $0, 1, \ldots, [-\lambda_i/(\sigma^2 - 1))$
$+1$.[5] Thus, the random count variable now has a theoretical maximum.

---

[5]If a is an integer, the function int[a] yields $a - 1$; otherwise, it returns the largest whole
number. For example, int[2.9] = 2, int[3] = 2, and int[3.1] = 3. Johnson and Kotz (1969,
p. 41) defined this function with the more usual int(a), producing a distribution quite incon-
venient for representation as a likelihood because the last term in the denominator is exactly
zero if $-\lambda_i/(\sigma^2 - 1)$ is an integer.

This is because in this situation of negative contagion, the initial events reduce the probability of future events and hence also reduce the maximum number of events that could occur in the period. Note that this maximum is a function of the parameters and is thus estimated rather than assumed. When σ^2 approaches one (indicating Poisson dispersion), this maximum approaches infinity just as in the Poisson distribution. The smaller σ^2 gets, the smaller becomes the maximum, although the maximum count is always larger than the expected value. Note also that $D_i = 1$, and the continuous parameter binomial reduces to the simpler binomial distribution if $(-\lambda_i)/(\sigma^2 - 1)$ is an integer. To my knowledge, this distribution has never been used for an event count regression model or for any other purpose in the social sciences.

The log-likelihood function for this distribution is as follows:

$$
\ln L_{cpb}(\beta, \sigma^2 \mid y) = \sum_{i=1}^{n} \left\{ \ln \Gamma\left(\frac{-\lambda_i}{\sigma^2 - 1} + 1\right) \right.
$$
$$
- \ln \Gamma\left(\frac{-\lambda_i}{\sigma^2 - 1} - y_i + 1\right) + y_i \ln(1 - \sigma^2)
$$
$$
\left. + \left(\frac{-\lambda_i}{\sigma^2 - 1} - y_i\right) \ln(\sigma^2) - \ln(D_i) \right\} \qquad (15)
$$

Maximizing this log-likelihood function, then, yields optimal estimates for β and σ^2 in the presence of underdispersion due to negative contagion.[6] Just as with the negative binomial model for overdispersion, both $\hat{\beta}$ and $\hat{\sigma}^2$ can yield interesting information. Again β are the effect parameters, and σ^2, in this model, indicates the degree of underdispersion existing in the data and negative contagion giving rise to it.[7]

5. Modeling Event Counts with Unknown Dispersion

In this section I develop a new generalized event count estimator that enables researchers to model event counts with unknown degrees of under-, over-, or Poisson dispersion. With only the models described above, one would need to conduct a pretest of some sort to decide on the type of dispersion to assume and, thus, which of the above models to estimate. However, the multiple steps that pretest estimators require result in less efficient

[6]The continuous parameter binomial distribution is continuous in both λ_i and σ^2. However, the first derivative is discontinuous at a small, finite number of points. In theory, one needs to be careful in using nonlinear optimization procedures in this situation, preferably choosing ones that do not rely on gradient searches. In practice I find that the small discontinuities in the derivative function rarely cause a problem even with optimization methods that rely on gradients.

[7]Since this model meets all of Wald's (1949) assumptions, the maximum likelihood estimator is consistent. Furthermore, in those situations where $\lambda/(\sigma^2 - 1)$ is not an integer, the estimator is also asymptotically normal.

estimators than simultaneous procedures (see Judge et al., 1985). But, more important, the incorrect choice among the Poisson, negative binomial, and continuous parameter binomial alternatives can lead to inconsistency as well as inefficiency (Gourieroux, Montfort, and Trognon, 1984b).[8]

To construct this generalized estimator, I derive a new probability distribution with parameters λ_i and σ^2. The difference here is that σ^2 is no longer restricted to be in a narrow range, but can now take on any value greater than zero. Special cases of this distribution occur when the dispersion parameter, σ^2, falls into different ranges. Thus, when $0 < \sigma^2 < 1$, this generalized event count distribution produces the same probabilities as the continuous parameter binomial; when $\sigma^2 = 1$, probabilities are the same as the Poisson; and, when $\sigma^2 > 1$, probabilities are the same as the negative binomial. This new probability distribution offers smooth transitions between these cases.

The likelihood function may then be formed with this distribution and simultaneously maximized with respect to σ^2 and β (since we specify $E(Y_i) \equiv \lambda_i = \exp(x_i\beta)$ as usual). The virtue of this approach is that one need not choose among the three sets of assumptions about the process generating the counts ahead of time. Contagion, heterogeneity, negative contagion, independence, and other processes will produce an event count with a particular type of dispersion. From the event count data σ^2 may then be estimated in a single step simultaneously with the effect parameters, β. Since no additional parameters are introduced in this new distribution, and since it reduces to the three distributions discussed above in special cases (i.e., ranges of σ^2), *this method actually offers something for nothing*—that is a reduced chance for inconsistency with no cost in additional assumptions. The remainder of this section derives this new estimator; the next section demonstrates its rather dramatic advantages in two empirical applications.

To derive this new single probability distribution for the first time, I build on a result from theoretical statistics due to Leo Katz (1945, 1965). Katz showed that from the following very simple "bilinear recurrence relationship," one could calculate certain probabilities from the binomial, Poisson, and Pascal distributions:

$$\frac{f_k(y_i+1 \mid \theta_i, \gamma_i)}{f_k(y_i \mid \theta_i, \gamma_i)} = \frac{\theta_i + \gamma_i y_i}{y_i + 1}, \tag{16}$$

---

[8]If one had to use one of these models, but the type of dispersion were unknown, the best choice is probably the Poisson, or the negative binomial with σ^2 set arbitrarily to some number greater than one, since they are consistent in the face of some types of variance misspecification (see Gourieroux, Montfort, and Trognon, 1984a). However, these estimates are still inefficient and the standard errors can be extremely inconsistent. Fortunately, with the introduction of the more general estimator in this section, one need not choose among these inferior alternatives.

for $y_i = 0, 1, 2, \ldots$ and $\theta_i + \gamma_{i}y_i \geq 0$. He noted that certain negative binomial probabilities (a generalization of the Pascal distribution) can also be calculated, but did not notice, as Johnson and Kotz (1969, pp. 37, 40–43) did, that the binomial can be generalized to what I call the continuous parameter binomial. The main use of Katz's (1945, 1965) result has been in theoretical statistics and thus has not been used in empirical research. However, in three steps I now derive a more general probability distribution from these early results that enables the creation of a more useful estimator.[9]

First, I reparameterize equation 16 in order to make it comparable to the notation in previous sections. Analysis reveals that the expected value and variance of a random variable Y_i that adheres to the relationship in equation 16 are as follows:

$$E(Y_i) \equiv \lambda_i = \frac{\theta_i}{1-\gamma_i} \tag{17}$$

$$V(Y_i) \equiv \lambda_i\sigma^2 = \frac{\theta_i}{(1-\gamma_i)^2}$$

Then solving for θ_i and γ_i yields

$$\gamma_i = 1 - \frac{1}{\sigma^2}, \qquad \theta_i = \frac{\lambda_i}{\sigma^2} \tag{18}$$

Finally, I substitute the right-hand sides of these equations into equation 16, replace y_i with $y_i - 1$, and rearrange:

$$f_{gec}(y_i \mid \lambda_i, \sigma^2) = \begin{cases} \left(\dfrac{\lambda_i + (\sigma^2 - 1)y_i}{\sigma^2 y_i}\right)f_{gec}(y_i - 1 \mid \lambda_i, \sigma^2) & \text{for } y_i = 1, 2, \ldots \\ \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

Since I call the distribution to be derived the generalized event count distribution, I use f_{gec} as its notation. The expected value and variance of the distribution implied in equation 19 is now consistent with the models in previous sections: $E(Y_i) = \lambda_i$ and $V(Y_i) = \lambda_i\sigma^2$.

Second, equation 19 is still in the form of a recurrence relationship, so I now put it in a form closer to more traditional probability distributions. To do this, note that the term $f_{gec}(y_i - 1 \mid \lambda_i, \sigma^2)$ on the right-hand side

---

[9]Gurland and Tripathi (1975) extend the Katz system to three and four parameter families. Lee (1986) uses the Katz system to develop score tests of the Poisson model against the over- and underdispersion alternatives that could be used as a pretest estimator; see also Cameron and Trivedi (1986, p. 41).

requires one to recursively substitute in values in order to calculate probabilities. For example,

$$
\begin{aligned}
\Pr(Y_i = 2 \mid \lambda_i, \sigma^2) &= f_{gec}(2 \mid \lambda_i, \sigma^2) \\
&= \left(\frac{\lambda_i + (\sigma^2 - 1)2}{\sigma^2 2} \right) f_{gec}(1 \mid \lambda_i, \sigma^2) \\
&= \left(\frac{\lambda_i + (\sigma^2 - 1)2}{\sigma^2 2} \right) \left(\frac{\lambda_i + (\sigma^2 - 1)1}{\sigma^2 1} \right) \\
&\quad \times f_{gec}(0 \mid \lambda_i, \sigma^2)
\end{aligned}
\tag{20}
$$

The last line of this equation still contains $\Pr(Y_i = 0 \mid \lambda_i, \sigma^2) \equiv f_{gec}(0 \mid \lambda_i, \sigma^2)$, so further analysis is required before probabilities can be assigned numerical values. By using the insight from equation 20, I eliminate most of the recursion by rewriting equation 19 as follows:

$$
f_{gec}(y_i \mid \lambda_i, \sigma^2) = f_{gec}(0 \mid \lambda_i, \sigma^2) \prod_{j=1}^{y_i} \left(\frac{\lambda_i + (\sigma^2 - 1)j}{\sigma^2 j} \right)
\tag{21}
$$

for $y_i = 1, 2, 3, \ldots$ and zero otherwise.

Finally, I solve the problem with $\Pr(Y_i = 0 \mid \lambda_i, \sigma^2) \equiv f_{gec}(0 \mid \lambda_i, \sigma^2)$ by using one of the basic axioms of probability:

$$
\sum_{m=0}^{\infty} f_{gec}(m \mid \lambda_i, \sigma^2) = 1
\tag{22}
$$

which basically means that something happens in period i with probability one, whether that "something" is zero events, one, or more. Using this axiom to solve for $f_{gec}(0 \mid \lambda_i, \sigma^2)$ in equation 21 leads to this expression:

$$
f_{gec}(0 \mid \lambda_i, \sigma^2) = \left[\sum_{m=1}^{\infty} \prod_{j=1}^{m} \left(\frac{\lambda_i + (\sigma^2 - 1)j}{\sigma^2 j} \right) \right]^{-1}
\tag{23}
$$

Then, using standard results on the convergence of infinite series leads to the following solution for the probability of zero events:

$$
f_{gec}(0 \mid \lambda_i, \sigma^2) = \begin{cases} e^{-\lambda_i} & \text{for } \sigma^2 = 1 \\ (\sigma^2)^{-\lambda_i/(\sigma^2 - 1)} & \text{for } \sigma^2 > 1 \\ (\sigma^2)^{-\lambda_i/(\sigma^2 - 1)} D_i^{-1} & \text{for } \sigma^2 < 1 \end{cases}
$$

It thus turns out that this expression cannot be written in one equation, but this causes no particular problems because the function is still continuous in both σ^2 and λ_i, since, for example,

$$\lim_{\sigma^2 \to 1} \left[(\sigma^2)^{-\lambda_i/(\sigma^2-1)} \right] = e^{-\lambda_i} \tag{25}$$

Finally, the full probability distribution can be written by collecting the above results in one expression:

$$\Pr(Y_i = y_i \mid \lambda_i, \sigma^2) = f_{gec}(y_i \mid \lambda_i, \sigma^2) \tag{26}$$

$$= \begin{cases} f_{gec}(0 \mid \lambda_i, \sigma^2) \prod_{j=1}^{y_i} \left[\dfrac{\lambda_i + (\sigma^2-1)(j-1)}{\sigma^2 j} \right] & \text{for} \quad y_i = 1, 2, \ldots \\ e^{-\lambda_i} & \text{for} \quad y_i = 0 \text{ and } \sigma^2 = 1 \\ (\sigma^2)^{-\lambda_i/(\sigma^2-1)} & \text{for} \quad y_i = 0 \text{ and } \sigma^2 > 1 \\ (\sigma^2)^{-\lambda_i/(\sigma^2-1)} D_i^{-1} & \text{for} \quad y_i = 0 \text{ and } 0 < \sigma^2 < 1 \\ & \text{and } y_i \leqslant \left[\dfrac{-\lambda_i}{\sigma^1-1} \right] + 1 \\ 0 & \text{otherwise} \end{cases}$$

For fixed λ_i, $0 < \sigma^2 < 1$ yields the same probability as the continuous parameter binomial; $\sigma^2 = 1$ yields probabilities identical to those under the Poisson distribution; and, $\sigma^2 > 1$ yields negative binomial probabilities. Thus, the separate parts notwithstanding, equation 26 is a single probability distribution. Due to equation 25, this distribution is also continuous in both σ^2 and λ_i.[10]

From this new probability distribution, a more general maximum likelihood estimator can be derived. The log-likelihood, reduced to sufficient statistics, takes a relatively simple form:

$$\ln L(\beta, \sigma^2 \mid y) = \sum_{i=1}^{n} \Bigg\{ C_i - y_i \ln(\sigma^2) + \sum_{j=1}^{y_i} \ln[\exp(x_i\beta) + (\sigma^2 - 1)(j - 1)] \Bigg\} \tag{27}$$

where

$$C_i = \begin{cases} -\exp(x_i\beta) & \text{for } \sigma^2 = 1 \\ -\exp(x_i\beta)\ln(\sigma^2)(\sigma^2 - 1)^{-1} & \text{for } \sigma^2 > 1 \\ -\exp(x_i\beta)\ln(\sigma^2)(\sigma^2 - 1)^{-1} - \ln(D_i) & \text{for } 0 < \sigma^2 < 1 \end{cases} \tag{28}$$

[10]If there were a single function of λ_i and σ^2 that closely approximated $(\sigma^2)^{-\lambda_i/(\sigma^2-1)}$ and $\exp(-\lambda_i)$, representing equation 26 as a single equation would be possible, but this is not necessary.

The maximum of this function gives the values of β and σ^2 that have the highest relative likelihood of having generated the data. In practice I find that numerical maximization algorithms converge relatively easily and quickly with most applications.[11]

This method produces estimates of β and σ^2 without specifying whether the data are over-, under-, or Poisson dispersed.[12] The estimates are identical to those which result from the *correct* choice among the Poisson, continuous parameter binomial, and negative binomial models. Of course, with this method one need not worry that the correct choice has indeed been made, since the range and value of σ^2 is estimated rather than assumed.

6. Applications: Veto Challenges and Superpower Conflict

I present two examples in this section, one with underdispersion and the other with overdispersion. For both examples I estimate the standard Poisson regression model and a generalized event count (GEC) model. Since the Poisson is a special case of the GEC (with $\sigma^2 = 1$), any difference between the estimates demonstrates the advantages of the GEC. Both examples point to the clear superiority of this new model.

Veto Challenges and Underdispersion

Consider the number of U.S. presidential vetoes that Congress challenges each year (1946–84). Some of these challenges successfully overturn presidential vetoes; others do not. However, each challenge is a significant event in executive-legislative relations and U.S. national politics.[13] For the purposes of this demonstration, I include three explanatory variables: election year (coded one for presidential election years and zero otherwise),

---

[11]Anyone with some facility with matrix algebra could easily and quickly program Gauss, or a similar general purpose maximum likelihood routine, to implement this estimator. As one reviewer put it, "Quite frequently we see some new technique but we all know that it will take us six months to get the thing up and running so we never bother. This is not the case here. . . . . Anyone can do this stuff at home in their spare time." For those not so inclined, I have written a computer program called COUNT to estimate these, and many other, models for event count data (see King, 1988, 1989a, 1989b, 1989c). The program is available from Aptech Systems, Inc., 26250 196th Place South East, Kent, Washington 98042; (206)631-6679.

[12]For $\sigma^2 \geq 1$, this estimator meets all standard regularity conditions and thus has all the usual desirable properties of maximum likelihood estimators. For $0 < \sigma^2 < 1$, the estimator is still consistent but is only asymptotically normal if $\lambda_i/(\sigma^2 - 1)$ is not an integer for all i, a highly probable situation.

[13]The data are taken from King and Ragsdale (1988). A congressional challenge is defined as any veto for which at least one house of Congress held a formal roll call vote attempting to overturn the veto. A roll call vote is usually held if the issue has some reasonable chance of passing, although many challenges are not successful.

Congress (coded as the percentage of both houses of Congress who are members of the president's party), and Approval (the average annual percentage of the public who approve of the president). Theory suggests that fewer congressional challenges should occur during election years, when the president has many members of his party in Congress and when public approval of the president is relatively high. Since the number of vetoes per year provides a theoretical (although nowhere near the empirical) maximum on annual veto challenges, I also include the natural log of the number of vetoes.[14]

The key question for present purposes is the underlying process that generates the annual number of congressional challenges. If one were to collect data on individual challenges, this process could be modeled in the systematic component of the statistical model (see Rohde and Simon, 1985), but in the present data set this within-year process remains unobserved. Nevertheless, one can still theorize and test hypotheses about this process. The basic Poisson model is appropriate if one challenge has no influence on subsequent challenges. Since challenges are major events in U.S. politics, independence is probably not a reasonable assumption. Indeed, since each congressional challenge involves a substantial amount of political organization and lobbying efforts among both the president's friends and enemies, individual challenges are likely to be spread out. Thus, negative contagion seems probable: one challenge reduces the probability of another challenge during the next interval of time. The aggregate consequence of this substantive political process is underdispersion, $\sigma^2 < 1$. I test this hypothesis below.

To demonstrate the effects of too narrowly specifying the statistical model, Table 1 presents both the standard Poisson and generalized event count (GEC) estimators. The estimated effect parameters for both models are similar. As expected, additional vetoes permit more challenges. Presidential election years, and those years where presidents are popular with

---

[14]In the standard Poisson regression model, Maddala (1983, p. 52) and King (1989c) provide alternative justifications for using the natural log of the maximum count as a control variable. In both cases the probability of an event is assumed small enough so that this upper bound does not effectively change the shape of the underlying distribution. Maddala assumes that the systematic component of the model is really $\lambda_i/N_i = \exp(x_i\beta)$, so that the maximum count, N_i, is divided out. This form then implies that $\lambda_i = \exp(x_i\beta + \gamma \ln N_i)$ with γ constrained to one. Maddala points out that estimating, rather than constraining, γ to one is also a reasonable approach. King derives the same result by retaining the $\lambda_i = \exp(x_i\beta)$ systematic component, but letting $t_i = N_i$ in equation 1. However one prefers to incorporate this assumption, the same likelihood function results. One can extend this logic to the generalized event count model either directly with Maddala's derivation or with a new distribution with King's derivation. The latter would require one to begin with equation 1 instead of equation 2 in deriving the generalized event count distribution. Again, the results are identical.

the public, have fewer challenges than other years (see King, 1989c: section 5.2, for a discussion of parameter interpretation).

Although generally in accord with what one might expect from theory, results from the Poisson model indicate that these effects are quite imprecisely estimated. Test statistics (estimate/SE) calculated from this table indicate very high probabilities that these data were drawn from a population where all the coefficients are in fact zero. Except for ln(Vetoes), the probability levels are all above 0.72. These results would probably cause most investigators to conclude that challenges are some kind of random process or at least independent of the exogenous variables, to drop these and add different variables, or just to give up the project entirely.

However, a more interesting and flexible stochastic model of the process generating congressional challenges to presidential vetoes yields very different substantive conclusions. Since σ^2 is precisely estimated to be substantially less than one in the generalized event count model, the underdispersion hypothesis is clearly confirmed. Either a t-test of the σ^2 estimate or a likelihood ratio test for the difference between the two models confirms the substantial improvement to be gained by moving to the more realistic GEC model. In addition, although the coefficients are similar to those obtained from the Poisson model, their standard errors are strikingly smaller. A comparison of the standard errors in the last column of Table 1 indicates that by allowing the model to be underdispersed, the resulting standard errors are between 5.5 and 10.23 times smaller than under the Poisson model. t-tests lead one to conclude that all the coefficients, except possibly that for the election year variable, are very precise and significantly different from zero.

TABLE 1
Models of Congressional Veto Challenges

Variable	Poisson Estimate	SE	GEC Estimate	SE	Poisson SE ÷ GEC SE
Constant	0.516	1.888	0.838	0.211	8.95
ln (Vetoes)	0.888	0.351	0.847	0.034	10.23
Election year	−0.050	0.577	−0.077	0.083	6.95
Congress	−0.009	0.026	−0.011	0.004	5.50
Approval	−0.005	0.022	−0.007	0.003	8.67
$\hat{\sigma}^2$			0.121	0.043	

Poisson log-likelihood = 605.11.

GEC log-likelihood = 621.35.

Whereas an analysis with only the Poisson model would have led most researchers to draw conclusions about the absence of empirical effects, the generalized event count model was able to extract significantly more information from the same data. Some of the improved information came from more precise estimates of the effect parameters. But additional information about the underlying process generating congressional challenges to presidential vetoes—negative contagion—was also made available with this method. As this example demonstrates, these points are not merely technical improvements; they permit one to extract considerably more information from existing data and to model interesting political science processes more creatively and appropriately.

Explaining Superpower Conflict

Does military spending by the United States and the Soviet Union deter or provoke conflict between the superpowers? In addition, how much of observable superpower conflict is merely tit-for-tat responses to the other superpower? These are fundamental questions in international relations research with obvious and critical implications for national policy. To study these questions, one needs a measure of superpower conflict. The most likely candidate for such a measure is the list of conflictual events occurring between the United States and the Soviet Union. Fortunately, scholars in international relations have put considerable intellectual and financial resources into the collection of massive international event count data sets such as this. More than 40 such collections appear in the ICPSR collection alone (see Vincent, 1983).

The specific data source used for this study is the conflict and peace data bank (COPDAB; see Azar, 1982). The dependent variable is the number of conflictual actions directed from one superpower toward the other for each year 1951–78 ($Conflict_t$). For explanatory variables, I use lagged measures of Soviet military expenditures ($SUmilitary_{t-1}$), and U.S. military expenditures ($USmilitary_{t-1}$), both in constant 1970 billions of U.S. dollars (see Ward, 1984, p. 311). I also include a dummy variable for the party of the U.S. president, coded zero for Democratic presidents and one for Republicans. Recall that the Poisson model assumes that no contagion exists within an observation. Thus, the only way to include the hypothesis of at least some tit-for-tat behavior is to include a lag of the dependent variable as an explanatory variable; the coefficient on this variable would be a rough measure of the degree of tit-for-tat behavior. Of course, this would be only a very rough indicator, excluding almost entirely those individual events that are specific responses to the other superpower's actions.

The empirical results for the Poisson model appear in the first part of Table 2. The results seem to indicate that both U.S. and Soviet military ex-

penditures deter superpower conflict to a modest degree. Presumably also due to deterrence, Republican presidents seem to preside over less super-power conflict than Democrats. These findings are strongly reinforced by the very small standard errors and correspondingly large t-statistics on all three coefficients. However, the final coefficient, on the variable represent-ing conflict between the superpowers for the year before, is nearly zero with a very small t-statistic.

Thus, if one had only the Poisson model with which to analyze this data set, the seemingly inescapable conclusion (based on the last coefficient) is that superpower conflict is not contagious. Tit-for-tat behavior is either non-existent or, at most, is a relatively minor feature of U.S.-Soviet politics. In-stead, international conflict between the superpowers is mostly a function of domestically generated factors like previous levels of military spending.

Remarkably different inferences emerge from an analysis of the same data with the GEC model. Estimates appear in the second part of Table 2. Four of the five coefficients have decreased somewhat. More striking is that the Poisson standard errors are only about 6 or 7 percent of the GEC stan-dard errors (see the last column of the table). The evidence from this im-proved model indicates that each of the effect parameters is not different from zero by conventional significance levels. Even the coefficient on the lagged value of conflict is essentially zero.

However, these results do not imply that superpower conflict is purely random. Instead, the estimate of σ^2 is very large, indicating that a substan-tial degree of contagion exists at the unobserved level of the individual events within each year. This parameter thus seems to be picking up some

TABLE 2
Models of U.S.-Soviet Conflict

Variable	Poisson Est.	SE	t-stat.	GEC Est.	SE	t-stat.	P. SE ÷ GEC SE
Constant	5.3827	0.0689		5.0472	1.0197		0.07
US military$_{t\text{-}1}$	−0.0087	0.0011	−8.01	−0.0039	0.0157	−0.25	0.07
SU military$_{t\text{-}1}$	−0.0070	0.0004	−20.03	−0.0061	0.0052	−1.17	0.07
President	−0.3271	0.0110	−29.80	−0.2526	0.1855	−1.36	0.06
Conflict$_{t\text{-}1}$	0.0001	0.0001	1.05	0.0001	0.0015	0.05	0.07
$\hat{\sigma}^2$				15.8493	5.4670	2.72[a]	

Poisson log-likelihood = 9268.76.

GEC log-likelihood = 9431.12.

[a]This is a test of the hypothesis that $\sigma^2 = 1$, calculated as $(15.8493\text{-}1)/5.4670 = 2.72$.

of the true nature of tit-for-tat behavior between the two superpowers. For example, the United States claims to have caught a Soviet spy and expels a half dozen members of the Soviet embassy. In response, the Soviets immediately expel a dozen members of the U.S. embassy in Moscow. This tit-for-tat conflictual behavior or specific reciprocity may continue for several more iterations until one side eventually stops. The whole incident may take place over the course of a week or perhaps two or three at the most. With only the Poisson model and a lag of the dependent variable from a year ago, this feature of superpower conflict could not be found.

Indeed, with only data aggregated up to annual counts of events, finding any evidence of contagion with any model of these data is remarkable. Fortunately, the unobserved processes occurring within each observation have some observable consequences on the aggregate level. Unlike the more commonly used Poisson model, the GEC model is able to extract this sort of critical substantive information. It also enables a researcher to extract this information even when the precise form and nature of the underlying process is in doubt or even the subject of the inquiry.

7. Concluding Remarks

This paper introduces a common notation and presentation for the standard Poisson model of event counts, the negative binomial overdispersed model, and a new continuous parameter binomial underdispersed event count model. Until now, one had to make an assumption about the unobserved process generating one's data and consequently choose from among the first two models. Even with all three models available, an incorrect choice could result in inefficiency and inconsistency in the parameter estimates and substantial inconsistency in the standard errors. The new generalized event count model introduced here still requires one to make assumptions, but it permits a much more flexible specification that enables a researcher to estimate, rather than assume, interesting features of the unobserved processes. Thus, the method provides a way to extract information that was previously not ascertainable. It also turns out that this estimator produces more statistically efficient and consistent empirical estimates and standard errors without *any* cost in additional assumptions. By essentially expanding the parameter space of the dispersion parameter to cover the Poisson, continuous parameter binomial, and negative binomial distributions, the single generalized event count probability distribution provides a more flexible stochastic model from which a superior estimator was derived.

The advantage of this new approach is illustrated in applications to congressional challenges to presidential vetoes and U.S.-Soviet conflict. The former example is severely underdispersed, probably resulting from strong negative contagion. The overdispersion in the latter example is probably

due to a large amount of tit-for-tat interactions between the superpowers. In both cases information extracted with this new model is much richer than with the more commonly used Poisson regression model.

The models introduced here are all fully parametric and estimable with maximum likelihood. This approach has the benefit of producing the most efficient estimators possible, given the veracity of the model. A separate literature on robust estimation of event count models is concerned with making weaker assumptions by trading off varying levels of statistical efficiency. Whether one should make the trade-off that robust estimators require (weak assumptions, less efficiency) or the ones full information methods require (stronger assumptions, maximum efficiency) is never obvious and generally should depend on the particular research situation. I do not review these robust approaches here because that theory is still in the development stage and would nevertheless require far more space than is available. One troubling feature of these estimators is that they often do not provide very good information on the underlying process and degree and type of dispersion; these underlying processes are often considered ancillary, if not nuisance parameters. As I have argued above, these underlying processes can be of substantial substantive interest. In addition, most of these estimators are not robust to underdispersion or the negative contagious process from which it is generated and require much more complicated multistage estimation procedures.[15]

REFERENCES

Allison, Paul. 1984. *Event history analysis.* Beverly Hills: Sage.
Azar, Edward E. 1982. *The codebook of the Conflict and Peace Data Bank (COPDAB).* College Park: Center for International Development, University of Maryland.
Bishop, Yvonne M. M., Stephen E. Fienberg, and Paul W. Holland. 1975. *Discrete multivariate analysis.* Cambridge: MIT Press.
Cameron, A. Colin, and Pravin K. Trivedi. 1986. Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics,* 1:29–53.

[15]See Huber (1967) and White (1980, 1982) on "pseudo-maximum likelihood" estimators in general (sometimes called "quasi-maximum likelihood; see Harvey, 1981) and Gourieroux, Monfort, and Trognon (1984a, 1984b) on "pseudo-maximum likelihood estimators in the linear exponential family" such as the Poisson. See McCullagh (1983), Wedderburn (1974), and McCullagh and Nelder (1983) on "maximum quasi-likelihood" estimators and Gourieroux, Monfort, and Trognon (1984a, 1984b) on "quasi-generalized pseudo-maximum likelihood" estimators. Cameron and Trivedi (1986) give a very good review, and Hampel et al. (1986) provide an excellent, albeit technical, discussion of robust statistics in general.

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

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

———. 1984b. Pseudo maximum likelihood methods: Applications to Poisson models. *Econometrica*, 52:701–20.

Greenwood, M., and G. U. Yule. 1920. An enquiry into the nature of frequency distributions of multiple happenings, with particular reference to the occurrence of multiple attacks of disease or repeated accidents. *Journal of the Royal Statistical Society, Series A*, 83:255–79.

Gurland, John, and Ram Tripathi. 1975. Estimation of parameters on some extensions of the Katz family of discrete distributions involving hypergeometric functions. In Ganapati P. Patil, Samuel Kotz, and J. K. Ord, eds., *Statistical Distributions in Scientific Work*. Vol. 1, *Models and Structures*. Dordrecht-Holland: D. Reidel.

Hampel, Frank R., Peter J. Rousseeuw, Elvezio M. Ronchetti, and Werner A. Stahel. 1986. *Robust statistics: The approach based on influence functions*. New York: Wiley.

Harvey, Andrew C. 1981. *The econometric analysis of time series*. Oxford: Philip Allan.

Hausman, Jerry, Bronwyn H. Hall, and Zvi Griliches. 1984. Econometric models for count data with an application to the patents-R&D relationship. *Econometrica*, 52:909–38.

Holden, Robert T. 1987. Time series analysis of a contagious process. *Journal of the American Statistical Association*, 82.400:1019–26.

Huber, Peter. 1967. The behavior of maximum likelihood estimates under nonstandard conditions. *Fifth Berkeley Symposium on Mathematical Statistics and Probability I*. Berkeley: University of California Press.

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

———. 1970. *Distributions in statistics: Continuous univariate distributions I*. 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.

Judge, George G., W. E. Griffiths, R. Carther Hill, Helmut Lütkepohl, and Tsoung-Chao Lee. 1985. *The theory and practice of econometrics*. 2nd ed. New York: Wiley.

Katz, Leo. 1945. Characteristics of frequency functions defined by first order difference equations. Ph.D diss., University of Michigan.

———. 1965. Unified treatment of a broad class of discrete probability distributions. In Ganapati P. Patil, ed., *Classical and contagious discrete distributions*. Calcutta: Statistical Publishing Society.

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

———. 1988. Statistical models for political science event counts: Bias in conventional procedures and evidence for the exponential Poisson regression model. *American Journal of Political Science*, 32:838–63.

———. 1989a. Event count models for international relations: Generalizations and applications. *International Studies Quarterly*, 33:123–147.

———. 1989b. A seemingly unrelated Poisson regression model. *Sociological Methods and Research*, 17:235–255.

———. 1989c. *Unifying political methodology: The likelihood theory of statistical inference*. New York: Cambridge University Press.

King, Gary, and Lyn Ragsdale. 1988. *The elusive executive: Discovering statistical patterns in the presidency*. Washington, DC: Congressional Quarterly Press.

Lee, L. F. 1986. Specification tests for Poisson regression models. *International Economic Review*, 27:689–706.

Maddala, G. S. 1983. *Limited-dependent and qualitative variables in econometrics*. New York: Cambridge University Press.

McCullagh, P. 1983. Quasi-likelihood functions. *Annals of statistics*, 11:59–67.

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

McGowan, Patrick J., and Robert M. Rood. 1975. Alliance behavior in balance of power systems: Applying a Poisson model to nineteenth-century Europe. *American Political Science Review*, 69:859–70.

Mullahy, John. 1986. Specification and testing of some modified count data models. *Journal of Econometrics*, 341–65.

Neyman, Jerzy. 1965. Certain chance mechanisms involving discrete distributions. In Ganapati P. Patil, ed., *Classical and contagious discrete distributions.* Calcutta: Statistical Publishing Society.

Patil, Ganapati P., and M. T. Boswell. 1975. Chance mechanisms for discrete distributions in scientific modeling. In Ganapati P. Patil, Samuel Kotz, and J. K. Ord, eds., *Statistical distributions in scientific work.* Vol. 2. Dordrecht, Holland: D. Reidel.

Prentice, R. L. 1986. Binary regression using an extended beta binomial distribution, with discussion of correlation induced by covariate measurement errors. *Journal of the American Statistical Association, Applications*, 81.394:321–27.

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.

Russett, Bruce M. 1971. An empirical typology of international military alliances. *Midwest Journal of Political Science*, 15:262–89.

Schoenberg, Ronald. 1985. Latent variables in the analysis of limited dependent variables. In Nancy B. Tuma, ed., *Sociological methodology, 1985.* San Francisco: Jossey-Bass.

Stoll, Richard J. 1984. The guns of November: Presidential reelections and the use of force, 1947–1982. *Journal of Conflict Resolution*, 28:231–46.

Thompson, H. R. 1954. A note on contagious distributions. *Biometrika*, 41:268–71.

Vincent, Jack. 1983. WEIS vs. COPDAB: Correspondence problems. *International Studies Quarterly*, 27:147–77.

Wald, Abraham. 1949. Note on the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics*, 20:595–601.

Ward, Michael Don. 1984. Differential paths to parity. *American Political Science Review*, 78:297–317.

Wedderburn, R. W. M. 1974. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61:439–47.

White, Halbert. 1980. A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica*, 48:817–38.

———. 1982. Maximum likelihood estimation of misspecified models. *Econometrica*, 50:1–25.