The Generalization in the Generalized Event Count Model, with Comments on Achen, Amato, and Londregan

Gary King and Curtis S. Signorino

Abstract

We use an analogy with the normal distribution and linear regression to demonstrate the need for the Generalized Event Count (GEC) model. We then show how the GEC provides a unified framework within which to understand a diversity of distributions used to model event counts, and how to express the model in one simple equation. Finally, we address the points made by Christopher Achen, Timothy Amato, and John Londregan. Amato's and Londregan's arguments are consistent with ours and provide additional interesting information and explanations. Unfortunately, the foundation on which Achen built his paper turns out to be incorrect, rendering all his novel claims about the GEC false (or in some cases irrelevant).

Introduction

We are grateful to our discussants for their interest in our work on statistical models for event counts, especially their extensive inquiries into the generalized event count (GEC) model. This symposium is especially

Littauer Center North Yard, Harvard University; E-mail: gking@harvard.edu, curt@latte.harvard.edu. Our thanks to Chris Achen, Jim Alt, Tim Amato, Neal Beck, John Londregan, and Walter Mebane for many helpful discussions. The computer program COUNT, which implements the event count models discussed in this symposium and several others, is available in two versions. The more flexible version is available as part of the maximum likelihood modules in Gauss (a commercial statistical package available from Aptech Systems, Inc., 23804 South East Kent-Kangley Road, Maple Valley WA 98038; E-mail: sales@aptech.com). An easy to use, standalone, public domain version is available from http://GKing.Harvard.Edu.

timely given the increasing number of political science applications of these models[1] and methodological studies of them in other disciplines.[2] The data in many fields of political science are often coded as counts of discrete events in fixed periods, a data type with some unusual statistical properties. As all symposium participants and increasing numbers of applied political scientists recognize, methods tailored especially for event count data can easily outperform those developed for continuous unbounded data, such as regression analysis. And, more importantly, these new methods have the potential to extract significant new information from event count data sets.

The participants of this symposium have made a number of valuable contributions. Timothy Amato explores the derivation of the GEC distribution in more detail than previously attempted in the political methodology literature. John Londregan elegantly proves several important properties of the GEC. And Christopher Achen encourages political scientists to examine the data generation processes underlying our probability models.

In order to address our discussants' points most clearly, we begin by elaborating the GEC's unified framework for analyzing event counts and by explaining how its advantages over existing techniques work. We also demonstrate that the GEC model can be expressed in a single equation, in a format parallel to other related models. Because Amato and Londregan present the mathematical structure of the GEC, and Londregan so clearly proves its statistical properties, we do not repeat these necessary, but complex, mathematical arguments. Instead, we demonstrate most of our points conceptually, wherever possible discussing statistical models for event counts using an analogy with the linear regression model—the model with which political scientists seem to be most familiar. Of course, because other unifying frameworks could exist, and other methods of analyzing event counts do exist, we also emphasize the types of data sets to which the GEC is best applied.

One of the distinguishing features of political methodology is that many claims in this field are verifiably true or false. In the case of this symposium, all factual claims by Timothy Amato and John Londregan are true. Timothy Amato also poses a variety of interesting queries about the GEC model and its interpretation, many of which John Londregan conclusively answers and the remainder of which we also address. Unfortunately, the foundation on which Christopher Achen built his paper turns out to be incorrect, rendering all his novel claims about the GEC false (or in some cases irrelevant). We also explain these points here.

[1] See, most recently, Canon 1993, Krause 1994, Martin 1992, Nixon 1991, and Wang et al. 1993.

[2] See Winkelmann 1994 and the citations therein.

By way of a roadmap, we start by defining the types of data generation processes discussed in this symposium. We next demonstrate the need for a GEC model, using a linear regression analogy. The probability structure of the GEC and its asymptotic and finite sample properties are then presented. Finally, we discuss alternative approaches and offer suggestions for future research. An appendix provides technical details.

Data Types

We begin by summarizing the well-known definitions of event count data (the subject of this symposium), grouped binary data (a related data type sometimes confused with event counts), and the data type for which regression analysis is appropriate, since we will frequently use it as an analogy. The definitions we give in this section are identical to those used throughout the statistics and social science literatures (see King 1989a, sec. 3.2, 5.5–5.9; Maddala 1983, 51; Tuma and Hannan 1984; and McCullagh and Nelder 1989, 102, 193). Each of these three classes of data generation processes forms outcome variables in regression-type analyses. We save discussion of particular members of each of these classes until later sections.

Event Counts. These data are nonnegative integers $(0, 1, 2, \ldots)$ that represent the number of times a specified event occurs within a fixed observation period. Events occur with a (possibly varying) unobserved *expected rate of event occurrence* defined from the beginning to the end of each observation period, only at the end of which is the number of events observed, counted, and recorded. Examples include the number of coups d'état attempts per country and the number nonviolent protests per year.[3]

One of the most fundamental features of event count data is that the variance of the count increases with the expected number of events. Since counts are bounded from below by zero, event count observations with low rates of event occurrence must have relatively small variances. Count data with high rates of event occurrence have no such restrictions and, as a result, in virtually all real data sets, have higher variance. As

[3] More detailed observation of the frequency of events during each observation period could also be coded in some examples for which specific models have been derived (see Tuma and Hannon 1984). In many applications, analysts often decide not to record this more detailed information as "event histories" because the greater likelihood of measurement error and the frequent absence of covariates at this level of analysis often make the extra information not worth the effort.

we will explain, one of the mistakes in Achen's analysis results from his ignoring this key feature of event count data.

The maximum possible number of events in a count is unlimited, although all real examples have some practical limits on how large the counts will get. For example, although it is possible in principle for Rhode Island to experience 200,000 suicides in a single year, counts this large are so unlikely that the possibility can be effectively ignored.

The fixed period in which events are counted need not be identical for each observation in the data set. For example, the number of coups d'état can be counted for each country for 1995 or for all the years since each became an independent nation. (If the observation period varies, this information should be recorded, as it can be a valuable part of most statistical models, about which more below.)

Grouped Binary Data. Data of this type begin with a fixed and known number of unobserved trials during each observation period. Each trial has two (mutually exclusive and exhaustive) possible outcomes, conventionally labeled "success" and "failure," but they could be called "Fred" and "Barney" or anything else as long as the binary nature of the unobserved variable is maintained. The ultimate observed variable is an aggregate summary of these unobserved individual binary variables—the fraction of all trials that are "successes" $(0, 1, \ldots, n$ where n is known). For the trials within each observation, the probability of success may be the same or variable and the outcomes may be independent or related. Examples include the number of elections out of the last five in which a survey respondent reports having voted or the fraction of presidential nominations rejected by the Senate. The number of trials may vary over observations as long as this number is known.

Grouped binary data are similar to event count data in that both can be coded as nonnegative integers, but they result from fundamentally different data generation processes and should not be confused. The key is that grouped binary data are generated by the outcomes of a series of known, but unobserved, binary trials. In contrast, event counts are generated without any trials but instead by an unobserved rate of event occurrence. One result is that the maximum number of successful trials in grouped binary data is equal to the observed number of trials (or binary variables), whereas event count data usually have no explicit maximum count ex ante. For example, the number of phone calls you received last week is an event count because it is a nonnegative integer generated by some unobserved rate of phone calling. In contrast, the fraction of those phone calls from students is grouped binary data, with each phone call representing one binary trial. As with event count data, statistical models have also been designed especially for grouped binary data (see King 1989a, sec. 5.5, 5.6).

In practice, grouped binary data with large numbers of trials and very small probabilities of "success" are statistically indistinguishable from event counts. Because, in this special case, the maximum count is uninformative, and because the formal statistical models used become mathematically indistinguishable, most methodologists prefer to analyze this special case of grouped binary data as event counts.

Normal data. The data category we have in mind here includes outcome variables that are approximately normally distributed. They are continuous, unbounded, interval-level variables that (conditional on some explanatory variables) are unimodal and approximately symmetric but do not have especially "fat tails."

Generally, variables in the social sciences do not fit in this category exactly. In practice, however, numerous variables are close enough that analysts choose to model them with techniques that are best suited to normal data, such as linear regression. For example, although district-level vote proportions (in contested elections and conditional on incumbency) are limited to the interval [0,1], they fit this data type well (Gelman and King 1994).

Why a Generalized Event Count Model?

The GEC model adds a feature to statistical models of event counts that has always been a crucial component of linear-normal regression models for continuous data. Indeed, it has been such an integral part of linear-normal models that it is virtually never mentioned in that more familiar context. In order to understand this aspect of the GEC for event counts, we first describe this feature in the linear-normal model.

An Analogy from Linear Regression

Define the linear-normal model as usual by letting the continuous variable Y_i, conditional on covariates X_i, be modeled as a normal distribution with mean $E(Y_i) = \mu_i = X_i\beta$ and constant variance $V(Y_i) = \sigma^2$ (we use bold for $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}^2$, since β and σ^2 have have different meanings in event count models):

$$\begin{aligned} Y_i &\sim \mathrm{N}(y_i|\mu_i, \boldsymbol{\sigma}^2) \\ &= \mathrm{N}(y_i|X_i\boldsymbol{\beta}, \boldsymbol{\sigma}^2) \\ &= \frac{1}{\sqrt{2\pi}\boldsymbol{\sigma}} e^{-(y_i - X_i\boldsymbol{\beta})^2/2\boldsymbol{\sigma}^2}. \end{aligned} \tag{1}$$

The parameters to be estimated (the unknowns in this equation) are the vector of effect parameters β and the variance σ^2, both of which are constant over the observations.

For pedagogical purposes, King (1989a, 60) defined a special case of the normal model, which he called the *stylized normal.* The only difference is that the variance of this linear regression model is always one ($\sigma^2 = 1$):

$$\begin{aligned} Y_i &\sim \mathrm{N}(y_i|X_i\beta, 1) \\ &= \mathrm{STN}(y_i|X_i\beta) \\ &= \frac{1}{\sqrt{2\pi}} e^{-(y_i - X_i\beta)^2/2}. \end{aligned} \tag{2}$$

That is, this distribution has only one unknown, β. It also has a variance that is constant over all observations, as usual with linear regression models, but with its variance fixed at one instead of left to estimation.

The stylized normal is obviously a restricted version of the normal, and it will be correct only in the sense that a stopped clock is sometimes right when you happen to consult it—by coincidence. But suppose the only model we were aware of for continuous data was the stylized normal.

Leaving aside heteroskedastic cases, which are ruled out by both the normal and stylized normal regression models, applying the stylized normal model can result in three possible outcomes: the true variance σ^2 can be too low, which we denote as *understylized*; exactly one, which we have already labeled *stylized*; and too high, which we call *overstylized.* Figure 1 demonstrates the consequences of applying the stylized normal model to over-, under-, and stylized normal data. For each panel, we defined X_i as including a constant term and a single explanatory variable set to 500 evenly spaced values between zero and one. Then values of Y_i were created by drawing random numbers from a normal distribution with mean X_i (i.e., $\beta_0 = 0$ and $\beta_1 = 1$) and variance $\sigma^2 = 0.1$ for the understylized case, $\sigma^2 = 1$ for the stylized, and $\sigma^2 = 10$ for the overstylized data set.

We fit the same stylized normal model separately to each data set. Each figure represents the resulting estimate of the effect parameters as a solid line. As can be seen, point estimates of the slopes and intercepts of these lines are very similar, indicating that estimates of β under the stylized normal model are robust to the degree of over- or understylization.

Also represented in the figure are dashed lines drawn at twice the stylized normal's variance above and below the solid line. Because the stylized normal distribution assumes that the variance is one, the lines in

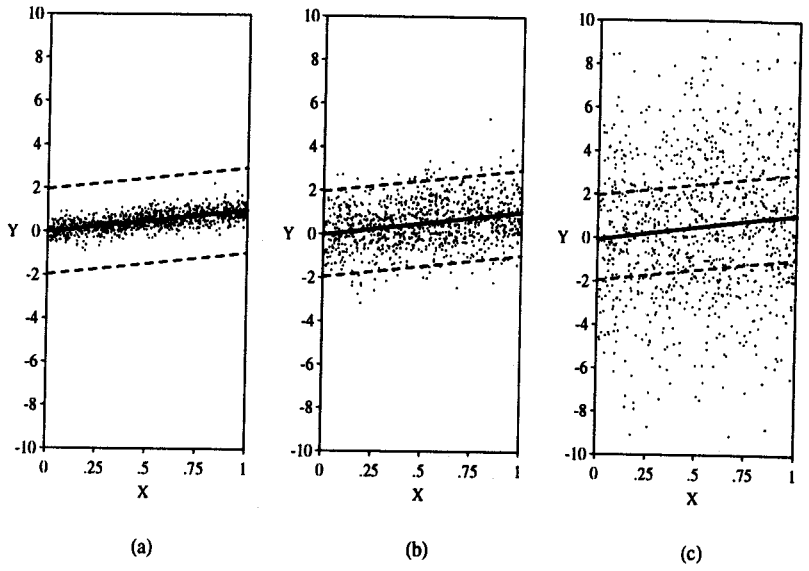

Fig. 1. Fitting a stylized normal model to different data types. The three data sets were drawn randomly from a normal model with (a) $\sigma^2 = 0.1$ (understylized), (b) $\sigma^2 = 1$ (stylized), and (c) $\sigma^2 = 10$ (overstylized). Note how the 95 percent confidence intervals, represented as dashed lines, are too large for understylized data and too small for overstylized data. Using a normal model instead, in order to estimate rather than assume σ^2, would have given the correct confidence intervals and would have fit each of the three cases. The purpose of this figure is to provide a simple analogy, under the more familiar normal model, to the consequences of applying the restrictive Poisson model to different data types. The analogous figure for event count data appears as figure 2.

the three graphs are each at the same fixed distance from the solid line.[4] The dashed lines are appropriately placed when about 5 percent of the points fall outside these lines, as occurs only for the stylized case. These results indicate that the stylized normal probability model is insensitive to the variance in the data and that the validity of the model will be greatly dependent on the degree of over- or understylization. Simply put, if σ^2 is not very close to one, the model will be wrong. For example, uncertainty estimates—such as standard errors of the coefficients or forecast confidence intervals—can be very far from the mark (way too large for understylized data and much too small for overstylized data).

[4] For simplicity, we ignore estimation variability in drawing these confidence intervals, which would make them look like parabolas with a dip in the middle when more data are nearby. With 500 observations drawn from the model, this effect is relatively minor in this situation.

In addition, many probability calculations from the model, such as the probability that $Y_i > 0.5$, will be wrong.

These problems with the stylized normal model are far from trivial. Even if one could rig together an alternative method of computing standard errors to evaluate its coefficient estimates, the model will always be wrong except for the rare case when $\sigma^2 = 1$. As such, even simple model checking procedures that involve evaluating any of a variety of the observable implications from the model become uninterpretable. In the end, there is usually little reason to choose a model that is so obviously rejected by the data.

The Advantage of the Generalized Event Count Model

The discussion of the linear normal regression model in the previous section provides a close analogy to a key problem in analyzing event count data under more restrictive models than the GEC. To summarize: the restrictive stylized normal for continuous data is to the restrictive Poisson model for event count data as the normal model for continuous data is to the GEC for event count data.

We begin by defining the GEC model. The GEC is a single probability distribution that includes a variety of other distributions as special cases, including every event count model discussed in this symposium. Some confusion on this point may have been caused by the original presentation of the GEC as a set of equations (King 1989b). Thus, we present the same distribution here in a more traditional format as a single equation (see the appendix for a derivation). Let the event count variable Y_i, conditional on explanatory variables X_i, be modeled as a generalized event count distribution with mean $E(Y_i) = \lambda_i = e^{X_i\beta}$ and variance $V(Y_i) = \lambda_i\sigma^2$.[5]

$$\begin{aligned} Y_i &\sim \text{GEC}(y_i|\lambda_i, \sigma^2) \\ &= \text{GEC}\left(y_i|e^{X_i\beta}, \sigma^2\right) \\ &= \frac{1}{y_i!}\left(\frac{e^{X_i\beta}}{\sigma^2}\right)^{\left(y_i, 1-\frac{1}{\sigma^2}\right)} \left[\sum_{j=0}^{y_i^{\max}} \frac{1}{j!}\left(\frac{e^{X_i\beta}}{\sigma^2}\right)^{\left(j, 1-\frac{1}{\sigma^2}\right)}\right]^{-1} \end{aligned} \tag{3}$$

where $y_i^{\max} = \infty$ for $\sigma^2 \geq 1$ and, letting $n_i = \lambda_i/(1-\sigma^2)$, $y_i^{\max} = \lceil n_i + 1 \rceil$ for $0 < \sigma^2 < 1$.[6]

[5] Winkelmann, Signorino, and King (1995) provide a small correction to these moments for the underdispersed case. Also, the $x^{(m,\delta)}$ notation is Signorino's (1995) δ-factorial; see the appendix.

[6] $\lceil x \rceil$ equals $x - 1$ for integer x and floor(x) for noninteger x.

The exponentiation function in the expected count keeps it greater than zero, which must be the case, given that the number of events can never drop below zero. This is unnecessary in the linear-normal model because continuous data are unbounded. Similarly, the variance is increasing with the mean, as with all reasonable models of event counts, by requiring a positive *dispersion parameter* $\sigma^2 > 0$. That is, σ^2 is not a variance term in this model but rather a positive factor that indicates how much the variance increases with the mean. Thus, this distribution has two unknowns, λ_i and σ^2, or, if explanatory variables are included, β and σ^2.

Just as we defined the stylized normal distribution by setting the normal distribution variance to one, we now define the Poisson distribution by setting the GEC distribution's dispersion parameter to one.

$$\begin{aligned} Y_i &\sim \text{GEC}\left(y_i|e^{X_i\beta}, 1\right) \\ &= \frac{1}{y_i!}\left(e^{X_i\beta}\right)^{(y_i,0)} \left[\sum_{j=0}^{\infty} \frac{1}{j!}\left(e^{X_i\beta}\right)^{(j,0)}\right]^{-1} \\ &= \frac{e^{e^{X_i\beta}}\left(e^{X_i\beta}\right)^{y_i}}{y_i!} \\ &= \text{Poisson}\left(y_i|e^{X_i\beta}\right). \end{aligned} \tag{4}$$

As will become apparent, the restrictive nature of the Poisson for event count data causes the same problems as the restrictive nature of the stylized normal does for continuous data.[7]

We now consider what happens when the Poisson model is applied to event count data other than generated by the Poisson. In our examination, we exclude data generation processes that have no reasonable analogies in real event count data. Thus, we do not examine homoskedastic cases or those for which the variance changes but does not increase with the mean. Of the range of remaining possibilities, we consider three: the true dispersion parameter σ^2 is smaller than the Poisson (and thus less than one), which is known as *underdispersed*; exactly one, which we have already labeled *Poisson*; and too high, which is called *overdispersed*. Poisson dispersed data arise when the rate of event occurrence is constant or otherwise unrelated to the number of events occurring during an observation. Overdispersed processes under the GEC occur

[7] Unlike the stylized normal, which we defined as a special case of the normal distribution, the Poisson distribution existed long before the more general GEC that includes it as a special case. However, although the history differs, the mathematical logic of the analogy does not.

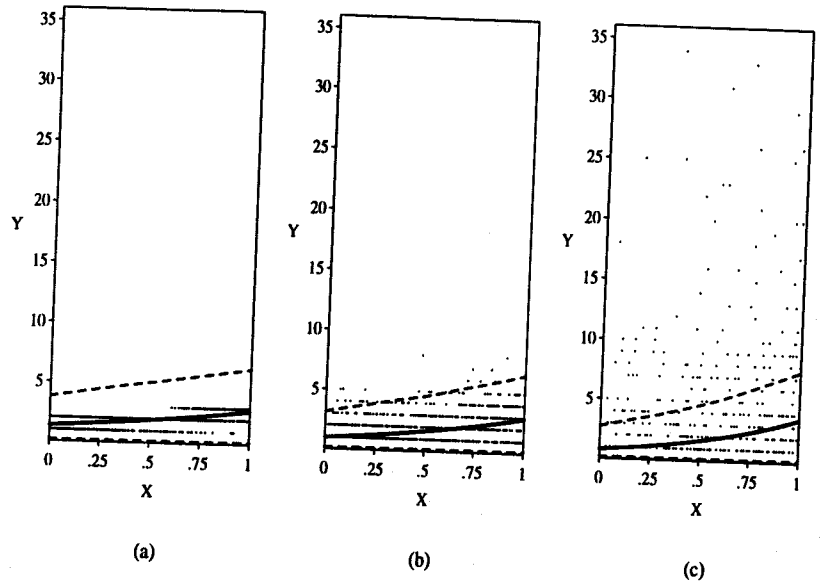

Fig. 2. Fitting a Poisson model to different data types. The data sets were drawn randomly from a GEC model with (a) $\sigma^2 = 0.1$ (underdispsersed), (b) $\sigma^2 = 1$ (Poisson dispersed), and (c) $\sigma^2 = 10$ (overdispersed). Note how the 95 percent confidence intervals, represented as dashed lines, are too large for underdispersed data and too small for overdispersed data. As we demonstrate in figure 3, applying a GEC model to these data instead, in order to estimate rather than assume σ^2, gives the correct confidence intervals and fits each of the three cases.

when the rate of event occurrence is positively correlated with the number of events (contagion) or varies randomly over time (heterogeneity). When increases in the rate of event occurrence depress the likelihood of future events (i.e., negative contagion), underdispersed count data result. Figure 2 completes the analogy to linear regression. For the three panels in this figure, we randomly generated data from the GEC model with the same mean but different dispersion parameters and fit the same restrictive Poisson regression model to it. In all three cases, we defined X_i as including a constant term and a single explanatory variable with values evenly spaced between zero and one and we set the mean to e^{X_i} (i.e., with $\beta_0 = 0$ and $\beta_1 = 1$). In panel (a), we drew data from the GEC with $\sigma^2 = 0.1$ as an example of the underdispersed case. Panel (b) data were drawn with $\sigma^2 = 1$, which is the Poisson; and panel (c) has data from the GEC with $\sigma^2 = 10$, which is our example of overdispersion.

The results of applying the restrictive Poisson model should by now be easy to anticipate: the point estimate of the mean function (and thus β) is reasonably well estimated in all three cases, but the model's variances, and thus uncertainty estimates, are not even close. In fact, most observable implications of the Poisson model are rejected in over- and underdispersed data. Just as with the stylized normal, there is little reason to choose a model like the Poisson when it is so clearly rejected by the data.

One possible saving grace for the Poisson would be if most data happened to be Poisson-dispersed in practice. Unfortunately, this is both rare in individual examples and, given the practice of most users, essentially impossible. The reason is that the "degree of dispersion" is always defined as *conditional* on the explanatory variables. As more explanatory variables are included, the true value of σ^2 (usually) declines from overdispersed, to the region of the Poisson, to underdispersed.

The result is that, if relatively few variables are included, and the data are (conditionally) overdispersed, the standard errors will be much too small and will cause the analyst to be wildly overconfident. It is not difficult to find examples in which standard errors are too small by a factor of 10 or even 100 or to generate artificial examples with much larger errors. In this situation, unknowing researchers will think they have discovered numerous new facts about the world, but few will be real. At the other end of the continuum, suppose the data are conditionally underdispersed, as occurs in practice if many control variables are included. If the Poisson regression model is applied in this situation, the standard errors will be much too large. In this case, the result will be that unwitting researchers will think no patterns exist in the data, potentially missing numerous important facts about the world.

The generalized event count model provides one possible solution to these problems, and it does so in the same way, at least conceptually, that the normal distribution solves the restrictive problems of the stylized normal. Rather than being forced to choose a value for σ^2 on the basis of little information prior to data analysis (or on the basis of some kind of more reasonable pre-test estimator), the GEC allows one to estimate β and σ^2 simultaneously. If data happen to be generated by a Poisson distribution, GEC estimates will indicate that $\sigma^2 \approx 1$ and the estimates and standard errors will be approximately the same as if you ran a Poisson in the first place. If instead the data are overdispersed, GEC estimates will indicate that $\sigma^2 > 1$, and other results will be identical to the results that would have come if a negative binomial model were run in the first place. Finally, if the data are underdispersed, the GEC extends the relationship between the Poisson and negative binomial to the region of the parameter space where $0 < \sigma^2 < 1$ and gives the same estimates as if the "continuous parameter binomial" distribution were

chosen. Because of this essential additional flexibility, the GEC model will fit the data better than the Poisson in the presence of under- or overdispersed data and essentially as well as the Poisson in the presence of Poisson dispersed data.[8]

The solution provided by the GEC can be seen in practice by following the usual practice of model fitting. With the GEC, including additional control variables (usually) makes the estimated value of the dispersion parameter σ^2 drop. In experiments of this sort, it is easy to get σ^2 to go from overdispersed, to approximately Poisson dispersed, to underdispersed, and the corresponding standard errors to similarly change, merely by including sufficient additional variables. The changes in σ^2 that occur for different sets of control variables provide powerful evidence in individual applications of the need for a generalized model like the GEC over a restrictive model like the Poisson. Indeed, since the Poisson nests within the GEC, the advantage of the GEC can be measured directly as the distance of σ^2 from one. If desired, one could also compute a hypothesis test to see whether $\sigma^2 = 1$, although since the GEC works as well as the Poisson in this situation there is little motivation for this test. An example of the superior fit of the GEC is provided in figure 3. For this figure, we use the three identical data sets presented in figure 2 and now substitute the Poisson regression model with the GEC regression model to estimate the coefficients, represented by the solid lines, and the 95 percent confidence interval as dashed lines around the solid line. As is obvious, the GEC fits all three types of event count data, both in mean and in variance around the mean. Note how the estimates of variability *correctly* and substantially drop from figure 2 to 3 in the underdispersed case and substantially increase in the overdispersed case.

Achen examines results in the literature that compare Poisson and GEC estimates for underdispersed and overdispersed data and, as expected, finds that the standard errors under the two models are "often strikingly different." He describes this difference as "worrisome," but instead of being correctly worried about researchers applying the restrictive Poisson model to over- or underdispersed data to which it does not fit, he incorrectly interprets this as a problem with the GEC. This mistake is logically equivalent to preferring the number 12 to the sample mean when estimating the average age in a population: although the sample mean is never worse than the number 12 as an estimator of the

[8] In theory, using the GEC when the Poisson is known to be the true model could add extra variability to estimates of σ^2. In practice, this possibly minor problem does not seem to arise. Of course, we are almost never in the position of knowing in advance that data are Poisson.

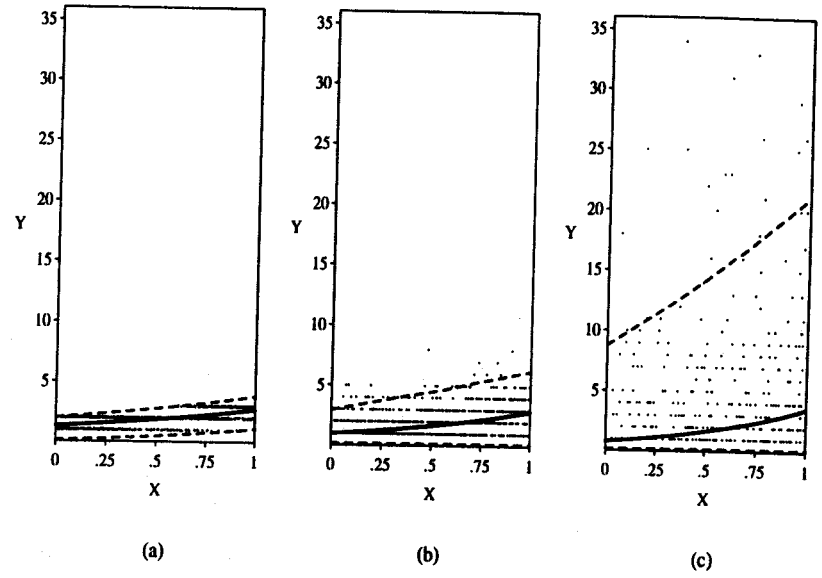

Fig. 3. Fitting the GEC model to different data types. The data sets represented in the three panels were drawn randomly from a GEC model with (a) $\sigma^2 = 0.1$ (underdispsersed), (b) $\sigma^2 = 1$ (Poisson dispersed), and (c) $\sigma^2 = 10$ (overdispersed). The solid line is the expected value, and the dashed lines are the 95 percent confidence intervals. These are the same data as those generated in figure 2, for which a Poisson model was estimated, but did not fit well for the under- and overdispersed cases. In contrast, the GEC easily fits all three data sets very well.

unknown quantity of interest, this error implies that we should be more suspicious about the sample mean as it gives estimates farther from 12.

This misunderstanding is exacerbated by Achen's discussion of an example in which, with underdispersed data, GEC standard errors happen to be half of the (incorrect) Poisson standard errors. He writes that "cutting standard errors in half is like quadrupling the size of one's data set . . . ." The mistake here is the implied claim that the Poisson standard errors to which he is comparing the GEC's are correct in the presence of underdispersed data and that the drop when moving to the GEC is only due to the supposed superior statistical efficiency of the GEC. However, as panel (a) in figure 2 demonstrates, his presumption that the Poisson model standard errors are correct in the presence of underdispersion is false. In addition, a comparison of figures 2 and 3 confirms that the advantage of the GEC in this case is not efficiency but rather that it, and not the Poisson, correctly fits the data. It would also be easy to see that this claim about the Poisson model is false by

using any of a variety of "robust" estimators of standard errors under the Poisson model such as White's "heteroskedasticity-consistent standard errors" provided as an option in COUNT.[9] Thus, moving to the GEC with underdispersed data will make your standard errors smaller, as they should be. Moving to the GEC with overdispersed data will increase your standard errors, as should also be the case.

Probability Models of Event Counts and Grouped Binary Trials

The rigorous proofs given by Londregan and definitions from Amato are sufficient to invalidate most of Achen's claims about probability distributions, including those about which distributions are special cases of others, those about the probabilistic structure of the GEC, and most of his table 2. But, even if his specific arguments were right, they would not be relevant empirically given the extremely special case to which his article is exclusively devoted. We discuss these points here.

Event Counts

We first list the probability distributions that are special cases of the GEC$(y_i|\lambda_i, \sigma^2)$ and the conditions, as indexed by ranges of σ^2, that produce them:

1. *Negative Binomial*, $\sigma^2 > 1$, the overdispersed case.
2. *Poisson*, $\sigma^2 = 1$.
3. *Continuous Parameter Binomial*, $0 < \sigma^2 < 1$, the underdispersed case. This special case itself reduces to an even more special case, the *Binomial*, when $n_i = \lambda_i/(1-\sigma^2)$ is an integer.

In the GEC, and therefore in each of these special cases, λ_i and σ^2 are unknown parameters to be estimated, with n_i having no substantive interpretation separate from λ_i and σ^2. Although the GEC distribution was derived from a *single* difference equation as its first principle (see the appendix, and King 1989b), and can itself be expressed in a *single* equation (see eq. 3), its more familiar special cases cause Amato to interpret it as a "'splicing' or 'joining' together of three separate distributions." Indeed, these distributions are joined, but they are joined under a more general common framework that encompass all of these familiar distributions as special cases. The GEC provides a single, consistent framework with which to understand this diversity of processes (and distributions) used to model event counts.

Thus, Achen's claim that the "GEC is a special case" of the binomial distribution is false. Exactly the reverse is true, as the binomial is a special case of the GEC.[10] To see this, start with the GEC and restrict σ^2 to the $(0,1)$ interval and fix n_i to an integer: the result is precisely the formula for the binomial distribution. In contrast, if we begin with the binomial, no specialization can produce the entire range of the GEC. Specializing the binomial cannot generate the full continuous parameter binomial (i.e., when $\sigma^2 < 1$ even if n_i is not an integer), the Poisson ($\sigma^2 = 1$), or the negative binomial ($\sigma^2 > 1$) ranges of the GEC. He also claims that the "GEC is a special case" of the negative binomial distribution (see his table 2), but this, too, is precisely backward: start with the GEC and restrict $\sigma^2 > 1$. The result is exactly the formula for the negative binomial. In contrast, no specialization of the negative binomial can produce the entire GEC, or its special cases of the continuous parameter binomial (for $\sigma^2 < 1$), or the Poisson ($\sigma^2 = 1$) distributions.

As a result, his claim that "The general rule is that the underdispersed GEC will produce the same fit as the conventional ... binomial model only when, apart from the intercept, none of the independent variables matter" is also false. The GEC will equal the binomial whenever $n_i = \lambda_i/(1-\sigma^2)$ is an integer and $\sigma^2 < 1$, no matter how much or little the independent variables matter.

Achen makes his broadest (incorrect) claim by quoting from an outdated version of Amato's paper, where Amato had hypothesized that the GEC was not a probability distribution because, he originally had thought, it did not meet the axioms of probability. If this were true, the GEC would indeed have been "fundamentally flawed." However, Amato has since concluded that his hypothesis was incorrect and he has corrected the statement in the final version of his paper, which appears as an article in this volume. Amato now writes: "The upshot of this

[9] Achen missed this point in part because he is under the mistaken impression that "heteroskedasticity-consistent" standard errors should not be used except in the presence of heteroskedasticity. He was perhaps mislead by the term *heteroskedasticity-consistent*, which seems to imply that it only corrects standard errors for heteroskedasticity when none is assumed, as we might want to do for the linear-normal model. In fact, this alternative method of computing standard errors is robust to the hessian not being equal to the outer product of the gradients. Heteroskedasticity can produce this in a linear model, as can assuming the wrong functional form for the heteroskedasticity that exists in the data. This includes situations such as applying the Poisson model to under- or overdispersed data, even though both data types and the Poisson model are all heteroskedastic in different ways. In fact, using Poisson estimates and these or other "robust" standard errors is a reasonable alternative to the GEC in the special case in which a researcher is only interested in inferences about the coefficients. This procedure should be treated with caution, of course, since with over- or underdispersion the Poisson model is not the data generation process.

[10] It is also not true that the GEC parameterization (when n_i is an integer and $\sigma^2 < 1$) is a special case of the binomial parameterization, as the two parameterizations are mathematically equivalent.

mathematical and probabilistic analysis is that one can write down a 'GEC' distribution from the difference equation ... that obeys the axiom constraining probabilities to the unit interval." Londregan ties up any potentially loose ends by providing a formal mathematical proof that the GEC meets the axioms of probability. (If the mathematics are confusing, it is easy to convince oneself of this point by searching for a counterexample: merely try to find a single value of both $\lambda_i > 0$ and $\sigma^2 > 0$ such that the computed probability falls outside the unit interval or for which the sum of the probabilities does not equal one.)

Event Counts versus Grouped Binary Data

Beyond this general incorrect claim, the remainder of Achen's paper is limited to the underdispersed case, which is by far the least common empirically. In fact, he limits this analysis considerably further to the knife-edged situation in which $n_i = \lambda_i/(1-\sigma^2)$ is exactly an integer, the binomial distribution. This is a remarkably narrow focus since the probability of it occurring in practice (given any continuous prior distribution on the parameters) is zero. This means that, even if all of his remaining claims were true, they would be irrelevant in practice. Although a zero probability event is not of much interest, it is a piece of the parameter space.[11] As such, it is reasonable to suppose that it might provide insight into some more general feature of the GEC. Unfortunately, as we now describe, he specializes even further than the binomial to a case that cannot occur in practice and has no bearing on models for event counts.

In its general mathematical form, the binomial distribution has the same two parameters as does the more general GEC, λ_i and σ^2, but where $0 < \sigma^2 < 1$ and $n_i = \lambda_i/(1-\sigma^2)$ is an integer. Achen's paper is devoted to a further specialization of this distribution where n_i is known ex ante for each and every observation ($i = 1, \ldots, N$)—what Amato calls the "known-n binomial." Unfortunately, n_i is *never* known in real event count data. To see this, we reformulate n_i into the mean and variance of the event count Y_i (a formulation Amato shows to be especially helpful):

$$n_i = \frac{E(Y_i)}{1 - \frac{V(Y_i)}{E(Y_i)}}. \tag{5}$$

[11] It sounds counterintuitive, but a zero probability event can occur in practice. For example, suppose you randomly draw a number from the uniform distribution on the interval between zero and two. Drawing a value of exactly one is possible even though it has zero probability of ever occurring.

Now try to think of an example of real event count data for which n_i, the ratio of the expected value to one minus the ratio of the variance to the expected value of the event count, is known and is exactly an integer. We think no examples of real event counts of this type exist and cannot imagine anything but the most contrived examples where n_i might be known. As a result, Achen's numerous comments about the known-n binomial, his numerical examples based on it, and his estimation procedures contain no relevant content for understanding or evaluating event count models. This is not a matter of definition or interpretation: all of his discussions based on these procedures are either false or irrelevant.

The known-n binomial model is sometimes used for grouped binary data, and this appears to be the root of Achen's confusion. In the context of the known-n binomial for grouped binary data, n_i is the number of trials parameter, which is known ex ante. In the context of the (unknown-n) binomial used as part of the GEC, n_i has no such interpretation. (Amato emphasizes this point explicitly by putting quotes around the phrase "number of trials" when labeling n_i in his discussions of event counts.) In fact, *event count data are not constructed from trials*, so there exists no unobserved binary variables or numbers of trials. For example, what would the unobserved binary variables be in analyzing the number of coup attempts per nation? Perhaps we might imagine that there existed a series of separate (independent) thoughts about coups, say all such thoughts among all senior officers in the armed forces, and that only some of these thoughts turned into actual coup attempts. More importantly, we would need to know a priori the number of such thoughts and that all the thoughts were independent of each other and had identical probabilities of turning into real coup attempts. For many (if not most) instances of event count data, making these types of assumptions seems specious at best.

In contrast, in order to conceptualize these data as event counts, we only need to imagine an underlying *expected rate of occurrence* of coup attempts. This expected rate may be variable over the nation's existence (it might be very high at first and during periods of hyperinflation). If the expected rate is constant, or conditionally independent of the observed count during an observation, this leads to the Poisson portion of the GEC. If the expected rate is heterogeneous or is correlated with the observed count, we have contagion, and thus overdispersion, leading to a model like the negative binomial part of the GEC. Finally, if a higher expected rate during an observation leads to fewer events, we have negative contagion and models like the continuous parameter binomial portion of the GEC.

If one's data are grouped binary rather than event counts, then the

models developed explicitly for these data should be chosen (such as those in King 1989a, sec. 5.5, 5.6) instead of those for event counts (such as those in King 1989a, sec. 5.7–5.9).[12]

As a result of these misunderstandings, a variety of Achen's other statements also turn out to be false. These include his claim that the known-n binomial "serves as the baseline for comparison with the GEC." In fact, the two are models of different data generation processes for completely different data types. Indeed, even if n_i were known, the binomial does not allow for Poisson or overdispersion, which are far more common empirically. He also writes: "the GEC depends on the implicit assumption that every binomial trial has the same probability of success across all observations and that the exogenous variables influence only the number of trials." In fact, event count data have no underlying trials and no corresponding "probability of success" for each one; they require no such assumption. To see that the ***probabilities*** of events computed from the GEC depend on the values of the explanatory variables, just compute

$$Pr(Y_i = y_i|\beta, \sigma^2) = \text{GEC}(y_i|\lambda_i, \sigma^2) \tag{6}$$

$$= \text{GEC}\left(y_i|e^{X_i\beta}, \sigma^2\right) \tag{7}$$

for any values of y_i, X_i, β, and σ^2. Note that if X_i has any effect (so that $\beta \neq 0$), the probability of any event always moves as X_i moves.

[12] In grouped binary data with n_i so large that event count models are indistinguishable from models for grouped binary data, the maximum count is known but provides little useful conditional probabilistic information. In almost all real data sets like this, although there is a theoretical maximum number of events per period, there sometimes exists a much smaller but unobserved practical maximum number of events that can realistically occur. In these situations, n_i is virtually always used to model the unobserved practical maximum; this makes these data essentially equivalent to event count data. Of course, even if you start with the known-n binomial in these cases, you need to decide what to do with the unknown n_i. If you guess n_i, but guess wrong, the same problems as those described in figure 2 will occur, since the known-n binomial includes no dispersion parameter. Of course, if you don't know n_i, the best procedure is to estimate it in some way, a logic that will quickly lead back to an approach like the GEC. In these situations, the theoretical maximum value of n_i may still be included in models of event counts, especially if it varies over the observations, by taking advantage of the form of all event count distributions that explicitly include a variable for the duration of each unit's observation period (instead of the usual procedure, which restricts them to being all the same). King (1989a, sec. 5.8) shows that you can achieve an equivalent effect by including the log of the maximum as an additional control variable.

Properties of the GEC

Asymptotic

The original claim was that the GEC is continuous in σ^2 and λ_i, consistent, and asymptotically normal except, when $\sigma^2 < 1$, in the zero probability event that n_i is an integer (King 1989b, 771, n. 7; 774; 776, n. 12). However, given the absence of a formal proof, Amato reasonably believes that this "warrants further investigation." Fortunately, Londregan provides exceptionally clear formal proofs that pin down some of these and other points.

Finite Sample

Because analytical results are often not available, finite sample properties of likelihood estimators are sometimes evaluated via simulated data sets, the advantage of which is that the true values of the quantities of interest are known. The usual procedure is that the data are randomly generated from the model, the estimator is applied, and the computed estimates are compared to the true values. Amato especially believes that more of these Monte Carlo experiments need to be conducted in this case. Figure 3 is an example of this type of analysis used for evaluating the GEC.

Achen studies the GEC with one 20-observation simulated data set, a simulation that forms the core of his article. Unfortunately, his claim that "all the explicit GEC assumptions are met" in these simulated data is false. This and other problems are easy to see by reexpressing his parameterization in the more common event count format.[13] Under his model, 10 observations are chosen so that $\lambda_1 = 0.5$ and $\sigma^2 = 0.95$, and the other 10 observations are chosen to match $\lambda_2 = 9.5$ and $\sigma^2 = 0.05$. It is easy to see that this data generation process does not meet the GEC assumptions since σ^2 is constant over observations in the GEC but varies between the two halves of this artificial data set. To correctly apply the GEC to data from this data generation process, we would merely apply the model twice, once to each data set. When this is done, the GEC assumptions are met, and the GEC model reproduces the true parameter values exactly. Alternatively, letting σ^2 vary between the two parts of

[13] To translate from Achen's parameterization to that more commonly used for event count models and used in the text, let $\lambda_i = np$ and $\sigma^2 = 1 - p$, but note that p and n do not have the interpretations he assigned to them for event count models. Under his interpretation, a known n_i and Poisson dispersion, which he refers to as "the canonical starting point for the study of event counts," would imply that, regardless of the expected count or rate of event occurrence, no events would ever occur!

the data set, as suggested by King (1989b, 766), would also exactly reproduce the true parameter values. As such, Achen's claim that "the distinction [between fixed and varying σ^2] is irrelevant" is incorrect.

Unfortunately, the problems with this artificial data set are more serious than merely being irrelevant for evaluating the GEC. They are also irrelevant for evaluating *any* reasonable event count model, since this data generation process omits one of the most distinctive features of virtually all real event count data—that they are heavily heteroskedastic, with variance increasing with the mean. Specifically, the mean increases substantially between the two parts of this data generation process, from $E(Y_i) = 0.5$ to $E(Y_i) = 9.5$, but the variance remains *constant* at $V(Y_i) = 0.475$. In Monte Carlo studies, data are sometimes generated from a probability model different from the one used for estimation; this is useful to test robustness to model assumptions in the face of data problems that are likely to arise in practice. (An example of this procedure appears in figure 2.) Because this homoskedastic event count data generation process generally does not occur in practice, it is of little relevance.[14]

Given the value of having correctly generated Monte Carlo results, we provide some here. In order to implement these, we modified the COUNT program. As written, COUNT had very occasionally taken an inordinately long time to converge. The reason is that COUNT uses an easy-to-implement maximization procedure based on numerical derivatives even though some (zero probability) points in the likelihood are undifferentiable. This is a minor problem in practice since, although the GEC is not globally concave, it appears to have no local minima. Thus, whenever iterations proceed without reaching convergence, a slightly less precise convergence criteria gives an answer sufficiently close to the global maximum (equivalent to, say, four rather than five digits of precision). Nevertheless, even a few simulations that do not converge quickly can make Monte Carlo studies take an inordinately long time to conduct. In order to avoid these problems, we implement a version of John Londregan's suggestion about an alternative maximization procedure.

[14] Even under Achen's misspecified version in his table 1, all of the relevant comparisons of coefficients under his alternative models differ by less than the standard error of their difference (where the standard error of the difference is computed as the square root of the sum of the two squared standard errors). Moreover, even if we did not know the data generation process, ample information in these data and reported results demonstrate that the model does not match the data generation process. For example, the differences in the variance matrix estimators Achen wonders about in footnote 12 is a valid test (and in this case confirmation) of misspecification and should have been used to correct the data generation process chosen or the model applied.

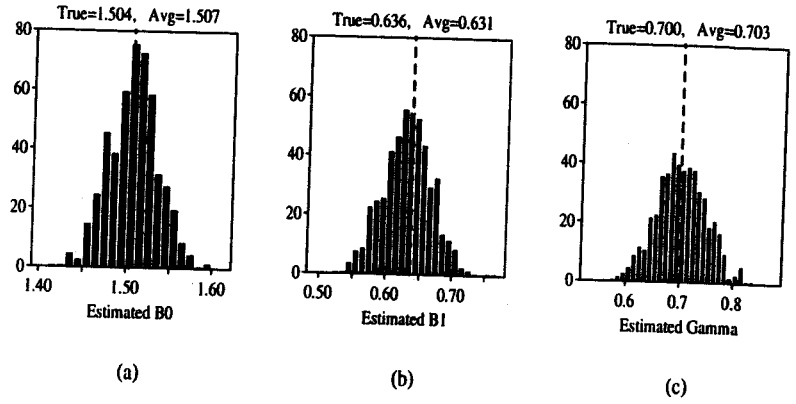

Fig. 4. Overdispersed data: Distributions of GEC parameter estimates from Monte Carlo runs. The figures show the true parameter values from which these overdispersed data were drawn (as dashed vertical lines) and the distribution of GEC estimates (as histograms) for (a) β_0, (b) β_1, and (c) $\gamma = \ln(\sigma^2)$. Note the approximately normal distribution of each set of estimates around its average, which is approximately the true value.

We use the usual derivative-based numerical optimization procedure until the search is fairly close to the maximum. We then use a grid search, when necessary, to pick out the parameter values that uniquely maximize the log-likelihood.[15] With this new procedure, we first drew 1,000 random data sets of 1,000 observations each from an overdispersed GEC model, with $E(Y_i) = \lambda_i = \exp(\beta_0 + \beta_1 X_i)$, $\beta_0 = 1.504$, $\beta_1 = 0.636$, and $\sigma^2 = 2$. We then estimated the parameters from each data set using the GEC model. Figure 4 presents these results. The true value of β_0, β_1, and $\gamma = \ln(\sigma^2)$ are indicated in the three panels with dashed vertical lines. The histograms plot the 1,000 estimated values. As is clear, the histograms are equal to the true values on average, with the spread distributed roughly normally around the true value.

We also randomly generated 1,000 underdispersed data sets, with

[15] Achen correctly points out that COUNT drops terms that do not depend on the parameters when reporting the value of the log-likelihood at the maximum. However, he also writes that this log-likelihood "is not the true value." In fact, because likelihood is a *relative* and not an absolute concept (King 1989a), there is no single "true" value of the log-likelihood. Add 75,261, or any other constant, to the log-likelihood and inferences do not change. Because dropping constant terms can save substantial computational time without any substantive consequence, COUNT, and most other programs that maximize likelihoods, drop these terms. The only thing to pay attention to is comparing different models that nest within one another but drop different constant terms.

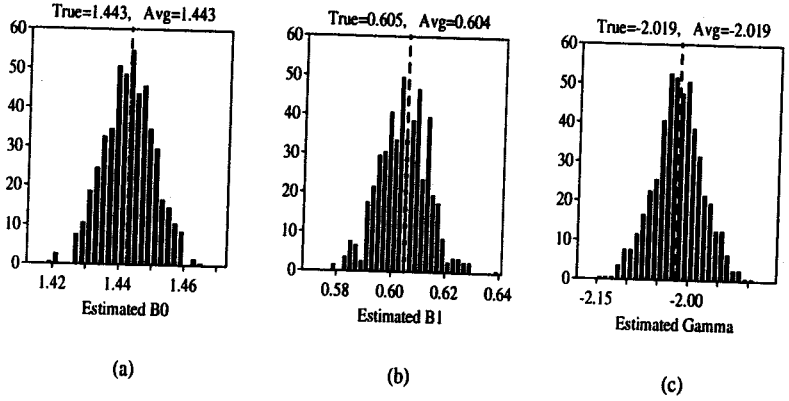

Fig. 5. Underdispersed data: Distributions of GEC parameter estimates from Monte Carlo runs. The figures show the true parameter values from which these underdispsersed data were drawn (as dashed vertical lines) and the distribution of GEC estimates (as histograms) for (a) β_0, (b) β_1, and (c) $\gamma = \ln(\sigma^2)$. Note the approximately normal distribution of each set of estimates around its average, which is approximately the true value.

$\lambda_i = \exp(\beta_0 + \beta_1 X_i)$, $\beta_0 = 1.399$, $\beta_1 = 0.636$, and $\sigma^2 = 0.1$. In order that these correspond to the expected value and variance, we applied the correction in Winkelmann, Signorino, and King (1995) to the true and estimated values, giving true values of $E(Y_i) = \exp(\beta_0^* + \beta_i^* X_i)$, $\beta_0^* = 1.443$, $\beta_1^* = 0.605$, and $\sigma_*^2 = 0.132$. We then ran the GEC model on the 1000 simulated data sets and recorded the estimates. Figure 5 presents these results in a parallel fashion to figure 4: the dashed vertical lines denote the true values, and the histograms summarize the results from the 1,000 simulations. As with the overdispersed case, the GEC model gives the right answer on average, with estimates approximately normally distributed around that estimate. The advantage of Monte Carlo analyses is that finite sample results can be obtained that often cannot be obtained via purely analytical means. The disadvantage is that they are necessarily limited to the specific set of parameter values chosen for the simulations. Thus, the Monte Carlo results presented here could be supplemented with many others, which we encourage future researchers to try. Analyses could be included for different parameter values, to evaluate the confidence intervals (as was done in figure 3), for different sample sizes, or for other reasonable data generation processes.

Alternative Approaches and Suggestions for Future Research

The GEC model allows for all ranges of dispersion, but as programmed it restricts the variance to be a positive scalar multiple of the mean, $V(Y_i) = \sigma^2 E(Y_i)$, so that the dispersion parameter σ^2, and thus the degree of under-, Poisson, or overdispersion, is constant for all observations. Amato is interested in the consequences of choosing alternative parameterizations of the variance function. Our view, consistent with that of McCullagh and Nelder (1989, 199–200), is that "even relatively substantial errors in the assumed functional form of Var(Y) generally have only a small effect on the conclusions," and so it is far more important to allow *some* kind of flexibility in deviating from the Poisson in the over- and underdispersed directions than to choose any particular form.[16] In all data sets we have studied, and all others with which we are familiar, the data do not contain sufficient information with which to distinguish among the various possible alternative variance functions, but obviously there is no guarantee that this will always be the case. We therefore briefly consider two possibilities here.

First, the constant σ^2 can be changed by allowing it to vary as a function of additional measured covariates Z_i in a parallel manner to λ_i. For example,

$$\lambda_i = e^{X_i\beta} \tag{8}$$

$$\sigma_i^2 = e^{Z_i\gamma} \tag{9}$$

where γ are variance function effect parameters to be estimated simultaneously with the mean function effect parameters β. This is a version of the strategy proposed in King (1989a, sec. 9.4; 1989c) and programmed in COUNT for the negative binomial case.

Another possibility is to change the basic form of the mean-variance relationship. For example, Winkelmann and Zimmermann (1991) propose generalizing the GEC variance function as

$$V(Y_i) = (\sigma^2 - 1)E(Y_i)^{k+1} + E(Y_i) \tag{10}$$

[16] Achen quotes twice from this same sentence in McCullagh and Nelder (1989), but he omits the piece that contains their central conclusion, which is quoted here. He also mentions the procedure suggested by McCullagh and Nelder to correct standard errors in the presence of overdispersed data when using the Poisson model. Unfortunately, although this procedure often pushes the standard errors in the right direction, even relatively simple Monte Carlo studies show that the procedure is far from optimal and usually does not solve the problem. Moreover, because the procedure is only intended to fix part of the problem, instead of providing a complete model, even when the procedure works, it does not result in a model that fits the data.

where k is estimated along with the other parameters. Special cases of this specification include $k = 0$, which gives the original scalar multiple relation; $k = 1$, which implies that the variance is a quadratic function of the mean; and a continuous range of other possibilities. Numerous other possibilities also exist that may be more appropriate in the context of specific data sets, and much future research awaits creative methodologists willing to explore them or develop new ones.

More generally, Amato "doubts that theories of politics are sufficiently well restricted to provide the requisite distributional assumptions to allow new distributions to be derived from first principles along the lines of the Greenwood Yule compounding result." Amato is obviously correct about theories of politics, and his point should make us especially aware of statistical models that are based on theories without a firm basis in reality or practical experience in real data. However, without methods derived especially for our data, political scientists are left using methods developed for studying insect longevity, chicken viruses, astronomical occurrences, and other nonpolitical applications. Some of these are applicable to our problems; others are not. Indeed, in many cases, we believe Amato's conclusion should be reversed. For example, without some approach like the GEC, or a pre-test estimation procedure, the only available models for event counts would require a theory of politics so restrictive that it would tell us in advance whether our data are over-, under-, or Poisson dispersed. With the GEC, or similar models, no such restrictive theories of politics are required. Political methodology is likely to make the most progress by understanding existing models and using them whenever appropriate and possible, but also by supplementing them with models better designed to meet the challenges of our field. Amato's warning is best answered by requiring new models to be developed for, applied to, and evaluated in real political data.

Finally, although improving the estimation of existing models is important, a much more interesting reason to develop new models especially for political data is to extract new information from these data so that we can learn more about the political and social world. In event count data, many such opportunities exist. For example, in the normal model, σ^2 is just the variance, whereas in the GEC σ^2 has the specific substantive interpretation of either heterogeneity, contagion, negative contagion, or Markov independence, each of which may have interesting substantive interpretations. Other models proposed for different types of event counts include hurdle models, seemingly unrelated count estimators, and multiple equation models (Mullahy 1986; King 1989c).

Many other opportunities for methodologists also exist in this field by focusing on precisely how their data are generated. For example, one of the largest supplies of event count data is in international relations. These data have special features that have not yet been included in any existing models. For example, international events data appear to have large amounts of systematic measurement error. In particular, the data include many observations for which no events were recorded but for which some probably did occur—as happens when, for example, coders are not available for some languages. In contrast, larger counts probably have less measurement error, or at least error that is more random. A model for this type of data, and other such processes, could make an important contribution to the analysis of international events data.

APPENDIX A. DERIVATION OF THE GENERALIZED EVENT COUNT PROBABILITY DISTRIBUTION

In this appendix, we provide an alternative derivation of the GEC, leading to the single-equation expression in equation 3 or the equivalent expression given originally in King 1989b. To simplify this result, we parameterize the distribution in terms of $\theta = \lambda/\sigma^2$ and $\gamma = 1 - 1/\sigma^2$. This causes no loss of generality because translating back to λ and σ^2 is straightforward.

We derive the GEC distribution starting from the Katz (1965) difference equation,

$$p(y|\theta,\gamma) = \left[\frac{\theta + \gamma(y-1)}{y}\right] p(y-1|\theta,\gamma). \tag{11}$$

In order for equation 11 to generate probability values, a constraint must be levied on y. Note that when $\gamma < 0$, we must have $\theta + \gamma(y-1) \geq 0$ for $p(y|\theta,\gamma) \geq 0$. For $\gamma \geq 0$, there is no constraint on y's maximum value. Therefore, define y's maximum value as

$$y_i^{\max} = \begin{cases} [1 - \frac{\theta}{\gamma}) & \gamma < 0 \\ \infty & \gamma \geq 0 \end{cases} \tag{12}$$

where $[x)$ equals $x - 1$ for integer x and $\text{floor}(x)$ for noninteger x.

Given $p(0)$, the values of $p(y|\theta,\gamma)$ can be stated as

$$
\begin{aligned}
P(1|\theta,\gamma) &= \theta p(0) \\
P(2|\theta,\gamma) &= \left(\frac{\theta+\gamma}{2}\right)\theta p(0) \\
P(3|\theta,\gamma) &= \left(\frac{\theta+2\gamma}{3}\right)\left(\frac{\theta+\gamma}{2}\right)\theta p(0) \\
&\vdots \\
p(y|\theta,\gamma) &= \frac{1}{y!}\left[\prod_{i=0}^{y-1}(\theta+\gamma i)\right]p(0). \qquad (13)
\end{aligned}
$$

Here, γ is the dispersion parameter and θ can be thought of as the arrival rate under Poisson dispersion. Examining equation 13, we see that when $\gamma = 0$, it has no effect on θ; when $\gamma < 0$, it *subtracts* from θ, yielding underdispersion; and, when $0 < \gamma < 1$, it *adds* to θ, yielding overdispersion.

We rewrite equation 13 as[17]

$$p(y|\theta,\gamma) = \frac{\theta^{(y,\gamma)}}{y!}p(0). \qquad (14)$$

Next, we derive an expression for $p(0)$ that allows equation 14 to generate probability values. Since the sum of the probabilities must

[17]Equation 14 uses the δ-factorial notation of Signorino (1995). For real x, nonnegative integer m, and real δ, the δ-factorial, $x^{(m,\delta)}$, is given by

$$x^{(m,\delta)} = \begin{cases} \prod_{i=0}^{m-1}(x+\delta i) = x(x+\delta)(x+2\delta)\cdots[x+\delta(m-1)] & m \geq 1 \\ 1 & m = 0 \end{cases}$$

Signorino shows that the δ-factorial subsumes falling factorials, rising factorials, and exponentials. Moreover, certain expressions commonly written using $\Gamma(\cdot)$ notation can contain singularities in the $\Gamma(\cdot)$ expressions, while the equivalent expression written using δ-factorial notation remains well defined.

equal one for any given combination of $\theta > 0$ and $\gamma < 1$, we have

$$
\begin{aligned}
p(0) &= 1 - \sum_{y=1}^{y_i^{\max}} p(y|\theta,\gamma) \\
&= 1 - p(0)\sum_{y=1}^{y_i^{\max}} \frac{\theta^{(y,\gamma)}}{y!} \\
&= \left[1 + \sum_{y=1}^{y_i^{\max}} \frac{\theta^{(y,\gamma)}}{y!}\right]^{-1} \\
&= \left[\sum_{y=0}^{y_i^{\max}} \frac{\theta^{(y,\gamma)}}{y!}\right]^{-1}. \qquad (15)
\end{aligned}
$$

Substituting this into equation 14, we get

$$p(y|\theta,\gamma) = \frac{\theta^{(y,\gamma)}}{y!}\left[\sum_{i=0}^{y_i^{\max}} \frac{\theta^{(i,\gamma)}}{i!}\right]^{-1}. \qquad (16)$$

It is straightforward to show that when $0 < \gamma < 1$, equation 15 simplifies to $(1-\gamma)^{\theta/\gamma}$ and that equation 16 reduces to a negative binomial distribution. Similarly, when $\gamma = 0$, equation 15 simplifies to $e^{-\theta}$ and Equation 16 reduces to a Poisson distribution. Finally, it can be shown that for $\gamma < 0$ and integer θ/γ, equation 16 simplifies to a binomial distribution.

REFERENCES

Canon, David T. 1993. "Sacrificial Lambs or Strategic Politicians: Political Amateurs in U.S. House Elections." *American Journal of Political Science* 37:1119–41.

Gelman, Andrew, and Gary King. 1994. "A Unified Method of Evaluating Electoral Systems and Redistricting Plans." *American Journal of Political Science* 38(2):514–54. (Replication data set: ICPSR s1054).

Katz, Leo. 1965. "Unified Treatment of a Broad Class of Discrete Probability Distributions" in Ganapati P. Patil, ed., *Classical and Contagious Discrete Distributions*, 175–82. Calcutta: Statistical Publishing Society.

King, Gary. 1989a. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. New York: Cambridge University Press.

King, Gary. 1989b. "Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator." *American Journal of Political Science* 33(3):762–84.

King, Gary. 1989c. "Event Count Models for International Relations: Generalizations and Applications." *International Studies Quarterly* 33(2):123–47.

Krause, George A. 1994. "Federal Reserve Policy Decision Making." *American Journal of Political Science* 38(1):124–44.

Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics.* New York: Cambridge University Press.

Martin, Lisa L. 1992. *Coercive Cooperation.* Princeton: Princeton University Press.

McCullagh, Peter, and John A. Nelder. 1989. *Generalized Linear Models.* 2nd Edition. London: Chapman and Hall.

Mullahy, John. 1986. "Specification and Testing of Some Modified Count Data Models." *Journal of Econometrics* 33:341–65.

Nixon, David. 1991. "Event Count Models for Supreme Court Dissents." *Political Methodologist* 4:11–14.

Signorino, Curtis S. 1995. "The Generalized δ-Factorial." Harvard University, mimeo.

Tuma, Nancy Brandon, and Michael T. Hannan. 1984. *Social Dynamics: Models and Methods.* Orlando: Academic Press.

Wang, T. Y., William J. Dixon, Edward N. Muller, and Mitchell A. Seligson. 1993. "Inequality and Political Violence Revisited." *American Political Science Review* 87(4):979–93.

Winkelmann, Rainer. 1994. *Count Data Models: Econometric Theory and an Application to Labor Mobility.* New York: Springer-Verlag.

Winkelmann, Rainer, Curtis S. Signorino, and Gary King. 1995. "A Correction for an Underdispersed Event Count Probability Distribution." *Political Analysis* 5:215–28.

Winkelmann, Rainer, and Klaus F. Zimmermann. 1991. "A New Approach for Modeling Economic Count Data." *Economics Letters* 37:139–143.