

*This article introduces a new estimator for the analysis of two contemporaneously correlated endogenous event count variables. This seemingly unrelated Poisson regression model (SUPREME) estimator combines the efficiencies created by single equation Poisson regression model estimators and insights from "seemingly unrelated" linear regression models.*

## A Seemingly Unrelated Poisson Regression Model

GARY KING  
*Harvard University*

### (1) INTRODUCTION

This article proposes an efficient method of estimating a seemingly unrelated Poisson regression model for the analysis of two correlated endogenous event count variables. This new estimator is intended as a solution to two problems with the analysis of event count data. This first problem is that, on the one hand, estimating equation-by-equation exponential Poisson models precludes improving efficiency by applying Zellner's (1962) "seemingly unrelated linear regression model" to the set of equations. On the other hand, the application of Zellner's full information maximum likelihood solution requires the assumption of Normally distributed disturbances and a linear or loglinear specification for each equation; with event count data, these implausible assumptions would mean a substantial sacrifice in

---

*AUTHOR'S NOTE: Thanks to Chris Achen, Bill Greene, Zvi Griliches, J. Scott Long, and Paul Zarowin for helpful comments on an earlier version of this article. Thanks also to the National Science Foundation for research support (NSF #SES-87-22715). An earlier version of this article was presented at the third annual meeting of the Political Science Methodology Group, Cambridge, Massachusetts, August 7-10, 1986.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 17 No. 3, February 1989 235-255  
© 1989 Sage Publications, Inc.

efficiency, bias, and even consistency (King, 1988). A second problem exists in performing cross-equation hypothesis tests with event count data, such as testing for the equality of coefficients in different equations; Zellner (1962) referred to another special case of these tests as "tests for aggregation bias." With such tests, one is presented with a choice among three unsatisfactory alternatives: (1) Do not do the hypothesis tests; (2) use equation-by-equation Poisson models, and assume that the covariance between the parameter estimates of the two equations is known a priori; or (3) use Zellner's model and do the hypothesis tests in the presence of bias, inefficiency, and inconsistency.<sup>1</sup>

This article proposes a joint Poisson regression estimator as a solution to these problems in the special case of two "seemingly unrelated" Poisson regressions. This estimator provides a full information maximum likelihood solution that is consistent and asymptotically more efficient than an equation-by-equation exponential Poisson model or Zellner's model applied to event count data. The larger the covariance between the two event counts, the greater the gain in efficiency. When the covariance is zero, the results are identical to the equation-by-equation estimators. The proposed method also permits reliable cross-equation hypothesis tests.

Section 2 provides background information on the linear, log-linear, and exponential Poisson models of univariate event count data and briefly reviews Zellner's seemingly unrelated linear regression model. Section 3 reviews the bivariate Poisson distribution in a form useful for further analysis. The proposed estimator is introduced in Section 4. Asymptotic efficiency advantages are demonstrated in Section 5, and Section 6 contains an application of the model to presidential vetoes of defense and social welfare legislation. Section 7 concludes.

## (2) BACKGROUND

Observation  $i$  ( $i = 1, \dots, n$ ) in event count data consists of the number of occurrences of an event in a fixed domain. The domain

for each observation may be time (a month, year, hour, or some appropriate interval) or space (a geographic unit, an individual, or others). This article is concerned with analyses with two contemporaneously correlated event count variables; observations thus vary over both time and space—a pooled time-series cross-sectional framework. Consider a few recent examples of univariate event counts: the number of consultations of a medical doctor for each survey respondent (Cameron and Trivedi, 1986), the number of triplets born in Norway in each half-decade (El-Sayyad, 1973), the number of presidential veto override attempts in the House and the number in the Senate per year (Rhode and Simon, 1985), the annual number of presidential appointments to the Supreme Court (King, 1987), the number of patents per firm (Hausman et al., 1984), the number of citizen-initiated and the number of support-related political activities engaged in and reported by Soviet emigres (Di Franceisco and Gitelman, 1984), the number of suicides per month, the number of spells of unemployment, and so on. There are numerous other examples from many disciplines.

Unfortunately, in many cases, event count data such as these are analyzed by expressing the expected value of each event count as a separate linear function of exogenous variables. This does not rule out negative, and therefore meaningless, fitted values; it results in substantial inefficiency; and, it almost surely uses the wrong functional form. In the single equation case there is little justification for the linear approach.

However, Zellner (1962) has shown that, by jointly estimating a set of linear regressions, efficiency can be improved over the equation-by-equation case. His model was not meant to analyze event counts, but it is sometimes used for this purpose. The key feature of this procedure is that it allows a fully efficient estimator even in the presence of contemporaneous correlations among the disturbances. The estimator is a “stacked” version of the linear model. Let

$$y_j = X_j\beta_j + \epsilon_j, \quad j = 1, \dots, M,$$

be the set of  $M$  equations, with  $k_j$  exogenous variables and  $N$  observations in each equation. Then, define  $K = \sum_{j=1}^M k_j$  and let  $\mathbf{y}$ ,  $\mathbf{X}$ ,  $\boldsymbol{\beta}$ , and  $\boldsymbol{\epsilon}$  be  $(MN \times 1)$ ,  $(MN \times K)$ ,  $(K \times 1)$ , and  $(MN \times 1)$  vertically stacked vectors of  $y_j$ ,  $X_j$ ,  $\beta_j$ , and  $\epsilon_j$ , respectively. Assume that  $E(\epsilon_j) = 0$  and  $E(\epsilon_i \epsilon_j') = \sigma_{ij} I_n$ . Hence, the variance of the joint disturbance vector is  $E(\boldsymbol{\epsilon} \boldsymbol{\epsilon}') = \boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes I$ , where  $\boldsymbol{\Sigma} = \{\sigma_{ij}\}$ . The estimator  $\hat{\boldsymbol{\Sigma}}$  is based on the equation-by-equation residuals,  $e_j = y_j - X_j \hat{\boldsymbol{\beta}}_j$ , and has elements  $\hat{\sigma}_{ij} = \epsilon_i \epsilon_j' / N$ . Finally, by letting  $\hat{\boldsymbol{\Omega}} = \hat{\boldsymbol{\Sigma}} \otimes I$  and assuming that  $\mathbf{X}$  has rank  $K$ , Zellner's seemingly unrelated linear regression model estimator can be written in a generalized least squares form:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X})^{-1} \hat{\boldsymbol{\Omega}}^{-1} \mathbf{y}.$$

Zellner has also shown that if  $\sigma_{ij} = 0$ , for all  $i \neq j$ , or if  $X_1 = X_2 = \dots = X_M$ , then the joint solution is identical to, and provides no improvement over, the equation-by-equation solution.

With correlated endogenous event count variables, Zellner's model will improve efficiency over the linear equation-by-equation model somewhat, but considerable inefficiency remains because the non-Normal distribution of the disturbances is not taken into account. The functional form and the risk of negative fitted values introduce further problems.

Another approach has been to take the natural log of  $\mathbf{y}$  and regress it on a vector of explanatory variables, either equation-by-equation or using Zellner's joint model. This partially corrects the heteroskedasticity and functional form problems. However, this approach is also problematic. Since the log of zero is undefined, the conditional expectation of  $\ln(y_i)$ , given a vector of exogenous variables  $\mathbf{x}'$  and given that  $y_i \sim \text{Poisson}(\theta_i)$ , is approximately equal to negative infinity, even in finite samples.<sup>2</sup> Moreover, adding an arbitrary small constant to  $y_i$  only complicates this problem by biasing the coefficients and making them highly sensitive to the choice of the constant; it is also not solved by using the positive Poisson distribution, or by replacing the logarithm with a square root transformation. Furthermore, Monte Carlo

experiments indicate that in finite samples this estimator is generally biased and between three and fourteen times less efficient than the Poisson regression estimator (King, 1988).

An alternative approach is to use the theoretically more appropriate exponential Poisson regression model (e.g., McCullagh and Nelder, 1983). Because a technique for *jointly* estimating a set of exponential Poisson regressions has not previously been available, opting for this model has also meant sacrificing certain efficiency gains. However, using this model does rule out negative fitted values and improves efficiency over the single equation linear and log-linear models. A special case of Nelder and Wedderburn's (1972) "Generalized Linear Model," with the following likelihood function, is generally applied in such instances:

$$L = \prod_{i=1}^n Pr(y_i | \theta_i) = \prod_{i=1}^n \frac{e^{-\theta_i} \theta_i^{y_i}}{y_i!}$$

$\theta_i$  is often assumed to be an exponential function of a vector of exogenous variables,  $x$ , both to ensure that fitted values never fall below zero, and for various substantive reasons (King, forthcoming a). Incorporating this assumption, the conditional expectation function can be written as:

$$E(y_i | X) \equiv \theta_i = e^{x_i' \beta}$$

The asymptotic variance of  $\hat{\beta}$  can be estimated by taking the inverse of the negative of the expected value of the second derivative of the log-likelihood function.<sup>3</sup>

$$V(\hat{\beta}) = \left( \sum_{i=1}^n x_i^2 e^{x_i' \hat{\beta}} \right)^{-1} \quad [1]$$

With a univariate model, this asymptotic variance is equal to the Cramer-Rao lower bound. However, the seemingly unrelated Poisson regression model will reduce this variance by simultaneously estimating two Poisson regression models. It eliminates the need to choose between two unsatisfactory alternatives—an inefficient equation-by-equation exponential Poisson model and an inconsistent seemingly unrelated log-linear regression estimator—and combines the advantages of both. This new estimator will also allow the testing of interesting cross-equation hypotheses. I begin with a description of the bivariate Poisson distribution on which the seemingly unrelated Poisson regression estimator will be based.

(3) THE BIVARIATE POISSON DISTRIBUTION

Campbell (1934; see also M'Kendrick, 1926) first derived the bivariate Poisson distribution by letting  $n \rightarrow \infty$  in a two-by-two contingency table, while constraining the marginal distributions to approach the Poisson limit. At the same time, the covariance approached a nonnegative limit. Marshall and Olkin (1985) provide a very useful review and also derive this distribution as the limit of a bivariate negative binomial. For present purposes, Holgate's (1964: 241; see also Johnson and Kotz, 1969: 297-298) approach and definition are most useful. My presentation in this section differs primarily in notation.

Let  $y_1^*$ ,  $y_2^*$ , and  $U$  be independently distributed Poisson variables with means  $\lambda_1$ ,  $\lambda_2$ , and  $\xi$ , respectively. Also let  $y_1 = y_1^* + U$  and  $y_2 = y_2^* + U$ . Then,  $y_1$  and  $y_2$  are each univariate Poisson with parameters  $\theta_1 = \lambda_1 + \xi$  and  $\theta_2 = \lambda_2 + \xi$ , respectively. The covariance of  $y_1$  and  $y_2$  is  $\xi$ . Together  $(y_1, y_2)$  are distributed bivariate Poisson:

$$Pr(y_1, y_2 | \theta_1, \theta_2, \xi) = e^{(\xi - \theta_1 - \theta_2)} \sum_{j=0}^{\min(y_1, y_2)} \frac{\xi^j}{j!} \frac{(\theta_1 - \xi)^{(y_1 - j)}}{(y_1 - j)!} \frac{(\theta_2 - \xi)^{(y_2 - j)}}{(y_2 - j)!}$$

Note that the bivariate Poisson and the bivariate Normal are among the very few distributions for which a zero covariance implies independence. This useful property, which has not been noted before for the bivariate Poisson, can be demonstrated by factoring this equation into the product of two marginal Poisson distributions, under the condition that  $\xi = 0$ :

$$\begin{aligned} Pr(y_1, y_2 | \theta_1, \theta_2, \xi = 0) &= e^{(-\theta_1 - \theta_2)} \sum_{j=0}^{\min(y_1, y_2)} \frac{0^j}{j!} \frac{(\theta_1 - \xi)^{(y_1 - j)}}{(y_1 - j)!} \frac{(\theta_2)^{(y_2 - j)}}{(y_2 - j)!} \\ &= \left( \frac{e^{-\theta_1} \theta_1^{y_1}}{y_1!} \right) \left( \frac{e^{-\theta_2} \theta_2^{y_2}}{y_2!} \right) \\ &= Pr(y_1 | \theta_1) Pr(y_2 | \theta_2) \end{aligned}$$

**(4) THE SEEMINGLY UNRELATED POISSON REGRESSION MODEL**

To define the proposed seemingly unrelated Poisson regression model (SUPREME), let  $\lambda_1$  and  $\lambda_2$  vary over  $i$  observations ( $i = 1, \dots, n$ ). Then assume

$$\begin{aligned} y_{1i}^* &\sim \text{Poisson}(\lambda_{1i}) \\ y_{2i}^* &\sim \text{Poisson}(\lambda_{2i}) \\ U &\sim \text{Poisson}(\xi) \end{aligned}$$

I further assume that  $y_{1i}^*$ ,  $y_{2i}^*$ , and  $U$  are independent at observation  $i$  and that, for observations  $i$  and  $j$  ( $i \neq j$ ), all three random variables are uncorrelated among themselves and each other. The observed dependent variables are then functions of these unobserved factors:

$$y_{1i} = y_{1i}^* + U$$

$$y_{2i} = y_{2i}^* + U$$

Thus  $y_{1i}$  and  $y_{2i}$  are distributed as bivariate Poisson with parameters

$$\theta_{1i} = \lambda_{1i} + \xi$$

$$\theta_{2i} = \lambda_{2i} + \xi$$

$$C(y_{1i}, y_{2i}) = \xi$$

for  $i = 1, \dots, n$ .

A consequence of the above assumptions is that  $y_{1i}$  and  $y_{1j}$  are uncorrelated and  $y_{2i}$  and  $y_{2j}$  are uncorrelated for all  $i \neq j$ . Thus, although there is no autocorrelation within each series, there is a constant contemporaneous covariance ( $\xi$ ) between the two variables.<sup>4</sup> Now let  $\theta_{1i}$  and  $\theta_{2i}$ , the expected values of their respective dependent variables at observation  $i$ , be exponential functions of separate linear combinations of exogenous variables and unknown parameters:

$$E(y_{1i} | X_{1i}) = \exp(x'_{1i} \beta_1)$$

$$E(y_{2i} | X_{2i}) = \exp(x'_{2i} \beta_2),$$

where  $x'_{1i}$  and  $x'_{2i}$  are vectors of  $k_1$  and  $k_2$  exogenous variables and  $\beta_1$  and  $\beta_2$  are coefficient vectors. The likelihood function of the SUPREME estimator,  $L^S$ , is then



$$L^S = \prod_{i=1}^n \Pr(y_{1i}, y_{2i} | \beta_1, \beta_2, \xi) \tag{2}$$

$$= \prod_{i=1}^n \exp(\xi - e^{x'_{1i}\beta_1} - e^{x'_{2i}\beta_2}) \sum_{j=0}^{\min(y_{1i}, y_{2i})} A_{ij}$$

where

$$A_{ij} = \frac{\xi^j}{j!} \frac{(e^{x'_{1i}\beta_1} - \xi)^{(y_{1i}-j)}}{(y_{1i}-j)!} \frac{(e^{x'_{2i}\beta_2} - \xi)^{(y_{2i}-j)}}{(y_{2i}-j)!}$$

The first and second derivatives of the log-likelihood function with respect to  $\beta_1$  can now be presented in a manner that makes later comparison with the derivatives of the univariate Poisson model convenient:

$$\frac{\partial \ln(L^S)}{\partial \beta_1} = - \sum_{i=1}^n x'_{1i} e^{x'_{1i}\beta_1} + \sum_{i=1}^n \left[ \frac{x'_{1i} e^{x'_{1i}\beta_1}}{e^{x'_{1i}\beta_1} - \xi} \frac{\sum_{j=0}^m A_{ij}(y_{1i}-j)}{\sum_{j=0}^m A_{ij}} \right]$$

$$\frac{\partial^2 \ln(L^S)}{\partial \beta_1 \partial \beta_1} = - \sum_{i=1}^n x_{1i}^2 e^{x'_{1i}\beta_1} \tag{3}$$

$$+ \sum_{i=1}^n \left\{ \frac{x_{1i}^2 e^{x'_{1i}\beta_1} \left[ e^{x'_{1i}\beta_1} \sum_{j=0}^m A_{ij}(y_{1i}-j) (\sum_{j=0}^m A_{ij} - 1) - \xi \sum_{j=0}^m A_{ij} \right] \sum_{j=0}^m A_{ij}(y_{1i}-j)}{(\sum_{j=0}^m A_{ij})^2 (e^{x'_{1i}\beta_1} - \xi)^2} \right\}$$

where  $m = \min(y_{1i}, y_{2i})$ .

### (5) THE GAIN IN EFFICIENCY

The asymptotic variance of the seemingly unrelated Poisson regression model can be estimated as usual for a maximum likelihood solution by taking the inverse of the expected value of the

negative of the second derivative in equation 3 (see King, forthcoming a). Let  $\psi$  be the parameter vector containing  $\beta_1$ ,  $\beta_2$ , and  $\xi$  vertically stacked in order, with  $k_1$ ,  $k_2$ , and 1 elements respectively. The variance of  $\hat{\psi}$  is then a  $(k_1 + k_2 + 1) \times (k_1 + k_2 + 1)$  symmetric matrix. Since this is a full information maximum likelihood solution, it is assured to have at least as small an asymptotic variance as the limited information equation-by-equation models. This variance matrix is shown here in partitioned form:

$$V(\hat{\psi}) = \left[ E \left( \frac{-\partial^2 \ln(L^S)}{\partial \psi \partial \psi} \right) \right]^{-1} \quad [4]$$

$$= \begin{bmatrix} E \left( \frac{-\partial^2 \ln(L^S)}{\partial \beta_1 \partial \beta_1} \right) & E \left( \frac{-\partial^2 \ln(L^S)}{\partial \beta_1 \partial \beta_2} \right) & E \left( \frac{-\partial^2 \ln(L^S)}{\partial \beta_1 \partial \xi} \right) \\ E \left( \frac{-\partial^2 \ln(L^S)}{\partial \beta_2 \partial \beta_1} \right) & E \left( \frac{-\partial^2 \ln(L^S)}{\partial \beta_2 \partial \beta_2} \right) & E \left( \frac{-\partial^2 \ln(L^S)}{\partial \beta_2 \partial \xi} \right) \\ E \left( \frac{-\partial^2 \ln(L^S)}{\partial \xi \partial \beta_1} \right) & E \left( \frac{-\partial^2 \ln(L^S)}{\partial \xi \partial \beta_2} \right) & E \left( \frac{-\partial^2 \ln(L^S)}{\partial \xi \partial \xi} \right) \end{bmatrix}$$

To demonstrate explicitly the gain in asymptotic efficiency that occurs when moving from the equation-by-equation Poisson model to the seemingly unrelated Poisson regression model, take the difference between two matrices: the variance of  $\hat{\beta}_1^S$  in the joint model [ $V(\hat{\beta}_1^S)$ , the 1, 1 element of equation 4] and the variance of the parameter estimate vector in the separate equation exponential Poisson model [ $V(\hat{\beta})$ , the matrix in equation 1]. If

$$V(\hat{\beta}_1^S) - V(\hat{\beta})$$

can be shown to be negative semidefinite, then the joint model would be proven to be asymptotically more efficient than the equation-by-equation Poisson model. Equivalently, the superior-

ity of the seemingly unrelated Poisson regression model can be demonstrated by showing that the following difference is positive semidefinite:<sup>5</sup>

$$\left\{ E \left( \frac{-\partial^2 \ln(L^S)}{\partial \beta_1 \partial \beta_1} \right) - E \left( \frac{-\partial^2 \ln(L)}{\partial \beta \partial \beta} \right) \right\}$$

$$= - \sum_{i=1}^n \left\{ \frac{x_{1i}^{\prime 2} e^{x_{1i}^{\prime} \beta_1} \left[ e^{x_{1i}^{\prime} \beta_1} \sum_{j=0}^m A_{ij} (y_{1i} - j) (\sum_{j=0}^m A_{ij} - 1) - \xi \sum_{j=0}^m A_{ij} \right] \sum_{j=0}^m A_{ij} (y_{1i} - j)}{(\sum_{j=0}^m A_{ij})^2 (e^{x_{1i}^{\prime} \beta_1} - \xi)^2} \right\}$$

The denominator and the two terms outside the square brackets in the numerator of this equation are positive. Thus, since the term in the square brackets is negative when  $\xi$  is not zero, the entire expression is positive semidefinite.

Furthermore, because of the position of  $\xi$  in the square brackets, the gain in efficiency will be larger when the covariance between the two dependent variables is larger. This result parallels that in the linear case (Zellner, 1962: 354), where a larger correlation also results in greater efficiency. Since this procedure can easily produce a test of the hypothesis that  $\xi$  is zero, one can readily test for the advantage of this joint technique over equation-by-equation estimation in empirical applications. This test is a relatively automatic by-product of the joint estimation procedure; there is no need also to run the equation-by-equation models.

One conceptual difference between the seemingly unrelated Poisson and seemingly unrelated linear regression models occurs when the independent variables from both equations are identical. In this situation, the latter "collapses" to yield single-equation least-squares estimators even if disturbance terms in different equations are correlated" (Zeller, 1962). No such limitation applies to the exponential Poisson estimator: Even when identical exogenous variables are used in both equations, a contemporaneous correlation among the disturbances will generally

yield a more efficient solution than equation-by-equation Poisson models.<sup>6</sup>

#### (6) AN APPLICATION TO PRESIDENTIAL VETOES

In *Congressional Government*, Woodrow Wilson (1885: 173) argued that the president's "power of veto . . . is, of course, beyond all comparison, his most formidable prerogative." Although this has been known for some time, presidency scholars have only recently begun to explain presidential vetoes systematically. To date, only highly aggregated analyses have been conducted. Empirical research in this area indicates that the use of the veto "varies according to the resources of the president and the character of the political environment" (Rhode and Simon, 1985: 411). Specifically, Rhode and Simon provide a model that explains the number of public bills vetoed per year as a linear function of public approval, the proportion of the president's party in both houses of Congress, international conflict, and the electoral cycle.<sup>7</sup>

Since presidential veto decision making is likely to differ when legislation is in defense policy versus social welfare policy, a more differentiated approach to explaining vetoes is warranted. There are other categories of congressional legislation than "guns and butter," but the well-established literature exploring the differences on other dimensions of these "two presidencies" makes these categories most likely to yield interesting findings (see Wildavsky, 1966; Sigleman, 1979; King, 1986).

For the purpose of this analysis, the number of presidential vetoes per year, 1946-1984, were individually coded into *social welfare* and *defense* policy categories (from *Presidential Vetoes, 1789-1976* and *Presidential Vetoes, 1977-1984*). Both the average annual public *approval* rating of the president and the average proportion of the president's party in both houses of *Congress* were included in the social welfare and the defense policy equations.<sup>8</sup> To measure the electoral cycle, the variable *cycle* was coded 1 if it was a presidential election year and 0 otherwise. Since

the frequency of social welfare legislation is far more likely to vary with the electoral cycle than defense legislation (Hibbing, 1984), *cycle* was not included in the defense veto equation. Finally, international *conflict* was coded as 1 for those periods when the U.S. military was engaged in combat during the entire legislative year and 0 for other years.<sup>9</sup> This variable was included only in the defense equation on the assumption that the decision to veto social welfare policy is independent of the existence of military conflict; preliminary analyses supported this assumption.<sup>10</sup>

Table 1 presents equation-by-equation and joint, linear, and log-linear estimates (with variances in parentheses). These estimates are problematic for the reasons indicated above, and the results provide some indication of these problems. The estimated linear models (see the first two columns of numbers) are implausible because when premultiplied by the explanatory variables they yield negative fitted values for 13% of the observations. This occurs for both the separate equation-by-equation results and for Zellner's joint estimator, although the magnitude of most of the negative values was slightly reduced in the latter case.

Since there were some years with no social welfare or military policy vetoes, a small constant was added, as usual, to the dependent variable to implement the log-linear procedure. The last five pairs of columns in Table 1 present estimates from the separate and jointly estimated log-linear models (with variances in parentheses). The different columns demonstrate the enormous sensitivity of the estimates to the value of the "arbitrary" constant. The joint parameter estimates are slightly less sensitive but even in this case one of the coefficients for  $\ln(y + 0.0001)$  is forty-five times what it is for  $\ln(y + 5)$ .<sup>11</sup> For the linear model and for the five log-linear models, the average efficiency of the joint procedure was between 2.96% and 6.48% greater than the equation-by-equation procedure. Improvement is relatively small because there are two identical explanatory variables in each equation.

The results indicate the implausibility of the linear and log-linear models, estimated either equation-by-equation or jointly as a set of seemingly unrelated linear regression equations. This



TABLE 2  
Poisson Models of Presidential Vetoes

Variable	Separate		Joint		Efficiency
	Estimate	Stan. Error	Estimate	Stan. Error	Advantage
<i>Social Welfare</i>					
Constant	1.182831	0.794691	1.337567	0.655517	1.469700
Approval	-0.005749	0.011832	-0.005551	0.009950	1.407565
Congress	-0.026928	0.011958	-0.028300	0.011916	1.006728
Cycle	0.599985	0.266175	0.349127	0.237643	1.254541
<i>Defense</i>					
Constant	0.271640	1.108794	0.390894	0.727171	2.325030
Approval	-0.049084	0.015362	-0.047385	0.014213	1.171908
Congress	0.049784	0.015264	0.054774	0.015133	1.019805
Conflict	-0.754597	0.452280	-0.947457	0.450889	1.006177
$\xi$			0.548920	0.181229	

is compounded by the analytical results indicating the relative inefficiency of these models compared to the Poisson models. It is of little interest, therefore, to interpret the value of any of these coefficients.

Next, consider the two exponential Poisson models. Table 2 presents the parameter estimates and approximate asymptotic standard errors. The estimate of  $\xi$  is statistically significant at the 0.001 level, indicating that the seemingly unrelated Poisson regression estimator is superior to the equation-by-equation estimators in these data. In addition, a likelihood ratio test, performed by adding the likelihoods of the separate models and comparing it to that for the joint model, indicates a similar improvement: The chi-square statistic is 34.56 with only one degree of freedom. The last column, which presents the ratio of the equation-by-equation Poisson regression variances to the seemingly unrelated Poisson regression variances, indicates that the joint model appears more efficient for every parameter estimate. The range of improvement is from 0.6% to 132.5%—33.3% on average. As an example of how large this improvement is, note that Zellner's (1962) seemingly unrelated regressions model improves efficiency by only about 0.20 times the equation-by-equation linear alternative. Note that finite sample standard errors will not necessarily be smaller for the SUPREME model

than the individual equation-by-equation models; the advantages of the former proven in Section 5 are all asymptotic, and standard errors are only asymptotic approximations in finite data. To the extent that the parameter estimates differ, concentration should generally be focused on the joint model.

Poisson models assume independence and homogeneity of the accumulating events within each period. When these assumptions are violated, parameter estimates are still consistent, but they are less efficient and the standard errors are biased (Gourieroux et al., 1984). For this reason, the equation-by-equation models are usually checked for the fit to the data by comparing them to a broader model based on the negative binomial or generalized event count distributions (King, forthcoming b). A check of the individual equations for this example indicates only very slight deviations from the Poisson assumptions (the data are about 15% overdispersed). In cases such as these, the minor efficiency one would theoretically gain with the negative binomial or generalized event count distributions would be lost because of having to estimate an additional parameter. More importantly, parameter estimates and standard errors across the different models in this case are almost identical. In theory we should also check the fit of the seemingly unrelated Poisson regression model estimator by comparing it to a more general joint model that allows for non-independence and heterogeneity. Developing a more general model such as this remains an important topic for research (see Hausman et al., 1984).

For a more specific interpretation of the coefficients of the exponential relationship, I take a linear approximation to it:

$$\frac{\partial E(y_i | X)}{\partial x_{ij}} = \beta_j e^{x_i \beta_i} = \beta_j \theta_i$$

Thus the linear effect of an explanatory variable on the expected number of events, holding constant the effects of the other explanatory variables, is the product of the corresponding coeffi-



cient and the expected value of the dependent variable. This expected value is of course unknown, and in any event varies over observations, so for interpretation one can replace it with the value of a typical observation, such as the empirical mean of the dependent variable. Alternatively, one could just think of the effect in proportion to the unobserved expected value.

The substantive results confirm my initial hypothesis that social welfare and military vetoes have different explanations. Previous research had used an aggregate model, incorrectly (albeit implicitly) assuming equal coefficients across these two equations, but even a cursory glance at this table indicates that this is not so. For example, the effect of presidential *approval* is essentially zero for social welfare vetoes. For defense policy, however, the coefficient indicates that a president who enjoys an approval rating increase of ten percentage points will decrease defense vetoes by about half of whatever is the expected number ( $[-0.047]10\theta = -0.47\theta$ ), other things being equal. Furthermore, a cross-equation hypothesis test for the equality of the two *approval* coefficients is rejected at the  $\alpha = 0.0041$  level, indicating that this difference was probably not due to random error.<sup>12</sup> The strong association between military matters and public approval of the president has long been noted (Mueller, 1975), but this is the first empirical demonstration that public approval has a substantially different effect on defense than social welfare policy. Rhode and Simon (1985: 412) found that approval decreased the aggregate number of vetoes, providing "supporting evidence for a proposition which has, in effect, remained untested for over twenty years." Table 1 lends credence to this proposition with respect to defense policy vetoes but not for social welfare vetoes.

As anticipated, a 10% increase in the number of members in Congress of the president's party decreased the frequency of presidential vetoes of social welfare legislation by about 28% of the expected number. Over four years with about average expected vetoing, this will do away with about one and a half presidential vetoes. In the case of military legislation, 10% more members of the president's party increase the number of vetoes by about 50% of the expected number. This difference does not

appear to be accounted for by random error, since the cross-equation hypothesis for the equality of coefficients is rejected at the  $\alpha < 0.0001$  level.<sup>13</sup> Whereas the opposition party in Congress, even as a majority, is not usually in a position to challenge the "bipartisan foreign policy" standard, members of the president's party, especially when in the majority, can usually dissent without such high political costs (King, 1986).

The remaining results were generally as expected. A presidential election year led to about a third more social welfare vetoes than the expected number, and the existence of military conflict decreased vetoing of defense legislation by about one veto a year.

### (7) CONCLUDING REMARKS

I have presented a full information maximum likelihood method for simultaneously estimating two seemingly unrelated Poisson regression models. This new estimator enables researchers analyzing event count data to take advantage of Zeller's (1962) insight that sets of equations estimated together can increase efficiency, thereby avoiding the major drawbacks associated with linear or log-linear models applied to event count dependent variables. It also allows meaningful tests of cross-equation hypotheses in the situation of contemporaneously correlated endogenous variables. These advantages were demonstrated both analytically and empirically.

### NOTES

1. Although Gallant's (1975) nonlinear model may be applied here, it does not take advantage of the non-Normal distribution of the disturbances.

2. For observations  $i$  ( $i = 1, \dots, n$ ),

$$E(\ln(y_i) | X) = \sum_{y_i=0}^{\infty} \ln(y_i) Pr(y_i | \theta_i) \approx -\infty.$$

Since this conditional expectation function does not exist, the estimator is also not meaningful:

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

3. Since this model is a special case of Nelder and Wedderburn's (1972), their iterative weighted least squares algorithm generally provides the quickest solution. In GLIM (Baker and Nelder, 1978), specifying a log link and Poisson error yields this model. See King (1988: Appendix 2) for other computer programs that will estimate this model. See also note 12.

4. The seemingly unrelated linear regression model makes almost identical assumptions.

5. Here I apply the result that if  $A$  and  $B$  are invertible matrices and  $A - B$  is negative semidefinite, then  $(A^{-1}) - (B^{-1})$  is positive semidefinite.

6. The variance-covariance matrix in equation 4 can easily be used for tests of cross-equation hypotheses. For example, if one were interested in whether the parameter on the first exogenous variable were the same for both equations

$$(\beta_{11}^S - \beta_{21}^S = 0),$$

one can calculate

$$v \equiv V(\hat{\beta}_{11}^S - \hat{\beta}_{21}^S) \equiv V(\hat{\beta}_{11}^S) + V(\hat{\beta}_{21}^S) - 2C(\hat{\beta}_{11}^S, \hat{\beta}_{21}^S)$$

where the first subscript refers to the equation and the second to the number of the coefficient in that equation. Then

$$\hat{\eta} \equiv (\hat{\beta}_{11}^S - \hat{\beta}_{21}^S)$$

is the point estimator,  $\eta^H$  the hypothesized value, and

$$\frac{\hat{\eta} - \eta^H}{\hat{v}^{1/2}}$$

forms an asymptotically standard Normal test statistic. The estimate of the covariance in this formula is a major contribution of this model, but it should also yield different parameter estimates and smaller variance estimates than in the equation-by-equation model. Considerably more complex cross-equation hypotheses can be similarly tested.

7. Except for the use of linear-Normal rather than an exponential-Poisson regression, Rhode and Simon conduct a relatively careful analysis, testing for and finding no evidence of autocorrelation or negative fitted values.

8. About every month, the Gallup polling organization asks a random sample of the American people, "Do you approve or disapprove of the way [the incumbent] is handling his job as president?" Preliminary analyses indicated that alternative measures of the Congress variable, such as dichotomous In party/Out party measure, did not perform as well as this one.

9. Other measures of conflict, such as the weighted number of conflictual events directed toward the United States, did not perform as well in preliminary analyses.

10. I have written a computer program called COUNT to estimate the SUPREME model, and many other models for event count data (see King, forthcoming a, forthcoming b, forthcoming c, 1988). The program is available from Aptech Systems, Inc., 26250 196th Place South East, Kent, WA 98042; 206-547-1733.

11. This sensitivity is also consistent with the analytical results given in King (1988). Some have argued that  $\ln(y + 0.5)$  should be used since  $E(\ln[y + 0.5]) - \ln(E[y]) \rightarrow 0$  as  $E(y) \rightarrow \infty$ . The problem with this logic is that the advantage of the Poisson models, and the need for the small constant to be added, are both greatest when  $E(y)$  is smaller.

12. According to the procedure described in note 6, the difference in the two coefficients is 0.0418, and the variance of this difference is  $0.000099 + 0.000202 - 2(0.000025) = 0.000251$ . The ratio of the difference to the square root of the variance is 2.640535, the value of an asymptotically standard Normal test statistic.

13. The difference between the two coefficients is 0.083074, and the variance is  $0.000142 + 0.000229 - 2(0.000008) = 0.000355$ . The value of the test statistic is then 4.4092.

## REFERENCES

- BAKER, R. J. and J. A. NELDER (1978) *The GLIM system*. Oxford: Numerical Algorithms Group.
- CAMERON, A. C. and P. K. TRIVEDI (1986) "Econometric models based on count data: comparisons and applications of some estimators and tests." *J. of Applied Econometrics* 1: 29-53.
- CAMPBELL, J. T. (1934) "The Poisson correlation function." *Proceedings of the Edinburgh Mathematical Society* 4: 18-24.
- DI FRANCEISCO, W. and Z. GITELMAN (1984) "Soviet political culture and 'covert participation' in policy implementation." *Amer. Pol. Sci. Rev.* 78(September): 603-621.
- EL-SAYYAD, G. M. (1973) "Bayesian and classical analysis of Poisson regression." *J. of the Royal Stat. Assn. Series B* 35(3)(Methodological): 445-451.
- GALLANT, A. R. (1975) "Seemingly unrelated nonlinear regressions." *J. of Econometrics* 3(February): 35-50.
- GOURIEROUX, C., A. MONFORT, and A. TROGNON (1984) "Pseudo maximum likelihood methods: theory." *Econometrica* 52: 681-700.
- HAUSMAN, J., B. H. HALL, and Z. GRILICHES (1984) "Econometrics models for count data with an application to the patents-RD relationship." *Econometrica* 52(July): 909-938.

- HIBBING, J. R. (1984) "Electoral pressures and transfer payment voting in the U.S. Congress." *J. of Politics* 46(August): 846-863.
- HOLGATE, P. (1964) "Estimation for the bivariate Poisson distribution." *Biometrika* 51: 241-245.
- JOHNSON, N. L. and S. KOTZ (1969) *Discrete Distributions*. New York: John Wiley.
- KING, G. (forthcoming a) *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. New York: Cambridge Univ. Press.
- KING, G. (forthcoming b) "Variance specification in event count models: from restrictive assumptions to a generalized estimator." *Amer. J. of Pol. Sci.*
- KING, G. (forthcoming c) "Event count models for international relations: generalizations and applications." *International Studies Quarterly*.
- KING, G. (1988) "Statistical models for political science event counts: bias in conventional procedures and evidence for the exponential Poisson regression model." *Amer. J. of Pol. Sci.* 32(August): 839-863.
- KING, G. (1987) "Presidential appointments to the Supreme Court: adding systematic explanation to probabilistic description." *Amer. Politics Q.* 15(July): 373-386.
- KING, G. (1986) "Political parties and foreign policy: a structuralist approach." *Pol. Psychology* 7(April): 83-101.
- MARSHALL, A. W. and I. OLKIN (1985) "A family of bivariate distributions generated by the bivariate Bernoulli distribution." *J. of the Amer. Stat. Assn.* 80(June): 332-338.
- McCULLAGH, P. and J. A. NELDER (1983) *Generalised Linear Models*. London: Chapman & Hall.
- M'KENDRICK, A. G. (1926) "Applications of mathematics to medical problems." *Proceedings of the Edinburgh Mathematical Society* 44: 98-130.
- MUELLER, J. E. (1975) *War, Presidents, and Public Opinion*. New York: John Wiley.
- NELDER and WEDDERBURN (1972) "Generalized linear models." *J. R. Statist. Soc. A.* 135: 370-384.
- Presidential Vetoes, 1789-1976 (1978) Washington, DC: Government Printing Office.
- Presidential Vetoes, 1977-1984 (1985) Washington, DC: Government Printing Office.
- RHODE, D. W. and D. M. SIMON (1985) "Presidential vetoes and congressional response: a study of institutional conflict." *Amer. J. of Pol. Sci.* 29(August): 397-427.
- SIGLEMAN, L. (1979) "A reassessment of the two presidencies thesis." *J. of Politics* 41: 1195-1205.
- WILDAVSKY, A. (1966) "The two presidencies." *Transaction IV*(December).
- WILSON, W. (1885) *Congressional Government*. Cleveland: World Publishing.
- ZELLNER, A. (1962) "An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias." *J. of the Amer. Stat. Assn.* 57: 348-368.

*Gary King is Associate Professor of Government at Harvard University. He is a frequent contributor to scholarly journals in the areas of quantitative methodology and American politics, is author of Unifying Political Methodology: The Likelihood Theory of Statistical Inference (forthcoming, Cambridge University Press) and coauthor of The Elusive Executive: Discovering Statistical Patterns in the Presidency (1988, Congressional Quarterly Press).*