The Political Methodologist

Newsletter of the Political Methodology Section
American Political Science Association
Volume 4, Number 2, Fall 1991

Editor: Charles H. Franklin, Washington University
and University of Wisconsin–Madison

Associate Editor: Larry Bartels, Princeton University

Contents

Notes from the Editor 1

Gary King: Calculating Standard Errors of Predicted Values based on Nonlinear Functional Forms 2

Donald Green: Maximum Likelihood for the Masses 5

David C. Nixon: Event Count Models for Supreme Court Dissents 11

A Maximum Likelihood Syllabus 14

Ron Schoenberg: The Gauss Corner: OPTMUM and MAXLIK Return Codes 16

Larry Bartels: Examination Questions 17

Henry E. Brady: Call for Papers: Ninth Political Methodology Conference 18

John R. Freeman: A Call from the New Editor of *Political Analysis* 18

Contents of *Political Analysis*, Volume 3 18

New Officers for Political Methodology Section 19

Philip Schrodt: Political Methodology Explained 19

Software News and Views 19

Notes From the Editor

Charles H. Franklin
Washington University and
University of Wisconsin–Madison

This issue is largely devoted to maximum likelihood estimation. For the vast majority of us, this is new ground. Should we venture into this new area or will we find that it is a desert, where all who wander are soon lost? To be blunt, is it worth our time to learn more about maximum likelihood?

The only real justification for investing the time (aside from simple intellectual curiosity) is that ML provides us with new tools to do better political science. As political scientists, our primary motivation is the analysis of politics. If a statistical technique allows us to do that better, then it is likely we will want to pay the price of admission. If it fails to improve our real work, then we'd rather continue as before.

So what purpose does statistics serve in political science and how does ML fit into this? The role of statistics is to give us a formal language for evidence, inference and hypothesis testing. This language is only as good as our ability to apply it to the theoretical questions which interest us. Thus we must shape our statistical models to fit our substantive concerns, not the other way around. In order to do this we need more flexibility than we have found in the traditional statistical work done in the discipline.

It is not so much a matter of tools not being available, but our ability and willingness to shape them to our needs. For example, we have treated heteroskedasticity as an annoyance for years, most often ignoring the possibility altogether or at most simply using GLS to get "better" standard errors. This is not a matter of inability to use the tools. We haven't paid much attention to this aspect of the data not because we don't have the tools but because we don't particularly care about it. We don't care because heteroskedasticity is not seen as embodying any interesting substantive hypotheses.

But suppose we consider what mechanism might drive heteroskedasticity, not as a statistical annoyance but as a substantively important phenomenon. For example, candidates are said to relish ambiguity. But what effect does ambiguity have on voter perceptions of candidate positions? Surely not through changing the (conditional) mean of perception. Rather ambiguity shows itself in the variance surrounding the (conditional) mean. In short, variation in ambiguity implies heteroskedasticity. Our substantive concern is how ambiguity affects voter uncertainty about candidate location and our statistical model must incorporate this substantive hypothesis. For a ML based solution to this problem, see Franklin, *APSR*, Dec. 1991.

While we can "cure" heteroskedasticity using least squares just as well as we can using ML, the latter offers us greater flexibility to look at heteroskedasticity in a new light. That is we can view it as something we want to model because it incorporates substantive meaning.

How should we model heteroskedasticity? ML methods have the great advantage of making it easy to do, rather than making it possible to do. We could do something like this with least squares, it simply turns out to be easier to do using ML. And we can then get good tests of our substantive results. This is an irresistible combination.

As with any statistical method, ML is no panacea. As with regression, structural equations and LISREL before it, ML will take time for us to come to grips with its advantages and its drawbacks. More importantly, we need to educate ourselves and our students in the proper use of this approach to estimation. It is important to realize that ML is a method of estimation, not a particular statistical model. Thus there are many models, including simple regressions, which can be estimated using ML. It is not the method of estimation which is important. It is the flexibility we gain in model specification and the ease of estimation which ML provides which makes it worth learning.

This raises the pedagogical issues that we need to face. If political methodology is to serve to advance the substantive work of the discipline, then it must be increasingly tied to the solutions of the kinds of substantive problems we face. ML is a very flexible tool that lets us do that better than we can with other approaches. This is its strength, and this is why we should learn about it and teach about it. How we go about doing that is a topic for another issue of *TPM*.

I am anxious to receive articles on the direction of methods requirements and offerings in departments around the country. We face a significant problem of teaching increasingly advanced techniques to graduate students who are as unprepared as ever. Articles addressing this subject would be welcome. The deadline for the next issue of *TPM* is March 15, 1992. Correspondence related to that issue should be sent to me at Washington University. —*CHF*

Calculating Standard Errors of Predicted Values based on Nonlinear Functional Forms

Gary King[1]
Harvard University

Introduction

Whenever we report predicted values, we should also report some measure of the uncertainty of these estimates. In the linear case, this is relatively simple, and the answer well-known, but with nonlinear models the answer may not be apparent. This short article shows how to make these calculations. I first present this for the familiar linear case, also reviewing the two forms of uncertainty in these estimates, and then show how to calculate these for any arbitrary function. An example appears last.

The Linear Case

Suppose we run a linear regression of y on X producing a vector of coefficients, $b = (X'X)^{-1}X'y$. Now think about setting the explanatory variables to new values (X^p) and making a prediction under this new situation about the next value of the random variable Y, which we will call Y^p. Y^p has N elements, which is the number of predicted values (not the number of observations).

The prediction Y^p has two forms of uncertainty associated with it: The first is that its expected value $E(Y^p)$ is estimated (with $\hat{y}^p = X^p b$) and thus depends on the estimated value b. This *estimation uncertainty* can be systematically reduced by increasing the sample size. The second is *fundamental variability* in the dependent variable around this expected value. Thus, even if we knew β (as in the case where we had an infinite number of observations), and were thus able to observe $E(Y^p) = \mu = X^p\beta$, we would not expect that μ would provide perfect predictions.

In the linear case, we can think about the components of the variance in Y^p simply by writing down the prediction as $Y^p = X^p b + \epsilon^p$ and calculating its variance:

$$\begin{aligned} V(Y^p) &= V(X^p b) + V(\epsilon^p) \\ &= X^p V(b) X^{p\prime} + \sigma^2 I \\ &= \sigma^2 X^p (X^{p\prime} X^p)^{-1} X^{p\prime} + \sigma^2 I \end{aligned} \tag{1}$$

where the two terms on the right side of this equation correspond to the estimation uncertainty and the fundamental variability, respectively. This is a simpler form of the proof in Johnston (1984: 199).

[1]Thanks to Neal Beck and Andrew Gelman.

To make the transition from the linear case to the more general case, we write this down and then analyze the full likelihood model (although in the linear case, one can make less restrictive assumptions). Thus, we assume a Normal stochastic component:

$$Y \sim N(\mu, \sigma^2 I) \tag{2}$$

and linear systematic component, $\mu = X\beta$. The likelihood function is then:

$$L(\beta, \sigma^2|y) = \prod_{i=1}^{n} N(X_i\beta, \sigma^2) \tag{3}$$

Under a properly specified likelihood model, such as this, the maximum likelihood estimate has an asymptotic Normal sampling distribution (see King, 1989, for a review of likelihood models and methods). In this case, the estimator b of β is distributed as follows:

$$\begin{aligned} b &\sim N(\beta, V(b)) \\ &= N\left(\beta, \sigma^2(X'X)^{-1}\right) \end{aligned} \tag{4}$$

We can now specify the distribution of the two quantities of interest in this analysis, $\hat{Y}^p$ and Y^p.

$$\hat{Y}^p \sim N(X^p\beta,\ X^pV(b)X^{p\prime}) \tag{5}$$

and

$$Y^p|\beta \sim N(X^p\beta,\ \sigma^2 I) \tag{6}$$

Since these distributions, representing estimation variability and fundamental variability respectively, are independent, we can combine them to derive the unconditional distribution for Y^p.

$$Y^p \sim N(X^p\beta,\ X^pV(b)X^{p\prime} + \sigma^2 I) \tag{7}$$

If we are really interested in knowing where the expected value is, then we would focus on the estimation variability in Equation 5. At other times we might be interested only in the variability around the expected value—variability in the world which beyond that due to estimation; this can be found in the fundamental variability of Equation 6. However, in most cases, we are interested in making predictions of where Y^p is likely to be on the basis of some specified explanatory variables X^p, as can be found in the total variability represented in Equation 7.

All Other Cases

To begin, I write down a general likelihood specification, with a stochastic component

$$Y \sim f(y|\theta) \tag{8}$$

and a (linear or nonlinear) systematic component, which for simplicity we assume is the expected value of the random variable:

$$E(Y) = \theta = g(X, \beta) = g(\beta) \tag{9}$$

The likelihood function is then written as usual

$$L(\beta|y) = \prod_{i=1}^{n} f(y_i|\theta) \tag{10}$$

The estimate of $E(Y)$ is the predicted value

$$\hat{y}^p = g(X^p, b) = g(b) \tag{11}$$

where b is the ML estimator of β. As in the linear case, the explanatory variable matrix X^p has one row for each predicted value, and so has dimensions $(N \times k)$.

Some examples of nonlinear functional forms include the exponential, logistic, or probit functions, respectively:

$$E(Y_1) = \exp(X\beta) \tag{12}$$

$$E(Y_2) = \frac{1}{1 + \exp(-X\beta)} \tag{13}$$

$$E(Y_3) = \Phi^{-1}(X\beta) \tag{14}$$

The fundamental variability can be calculated, conditional on knowing β, by the usual methods from probability theory:

$$V(Y) = E(Y^2) - E(Y)^2 \tag{15}$$

where

$$E(Y^2) = \int_{-\infty}^{\infty} y^2 f(y|\theta)dy \tag{16}$$

and

$$E(Y) = \int_{-\infty}^{\infty} yf(y|\theta)dy \tag{17}$$

In practice, one usually need not use Equations 15, 16, and 17 since the results for most popular distributions are widely available in books on probability theory. Thus, the fundamental variability is λ in the Poisson distribution, $\lambda\sigma^2$ for the negative binomial (depending on parameterization), and σ^2 for the Normal.

Calculating the estimation variability will usually require more effort, since these calculations are not as widely reported. Since expectations and variances are linear operators, a general method of calculating these is by calculating the linear approximation to the arbitrary linear function—the Taylor series. The Taylor series approximation of $\hat{y}^p = g(b)$ is as follows:

$$\hat{y}^p = g(b) = g(\beta) + g'(\beta)(b - \beta) + \cdots \tag{18}$$

where $g'(\beta)$ is the first derivative of the functional form $g(\beta)$ (from Equation 9) with respect to β. If there are k elements of β, and N predicted values, $g(\beta)$ is $(N \times 1)$ and $g'(\beta)$ is $(N \times k)$.

We now drop all but the first two terms in Equation 18 (making the equality in that equation an approximation), and apply the variance operator:

$$\begin{aligned} V(\hat{Y}^p) &\approx V[g(\beta)] + V[g'(\beta)(b-\beta)] \\ &= g'(\beta)V(b)g'(\beta)' \end{aligned} \tag{19}$$

(No covariances are necessary because each of the terms in the Taylor series approximation are independent.)

The matrix on the right side of Equation 19 is an $(N \times N)$ matrix, whereas $V(b)$ is $(k \times k)$. The $V(\hat{Y}^p)$ can be consistently estimated by substituting the estimated parameter vector and covariance matrix calculated by all standard ML routines for β and $V(b)$, respectively, in this equation. The standard errors of the elements of $\hat{Y}^p$ (based on estimation variability only) are the square roots of the diagonal elements of this matrix. Off diagonal elements of this matrix are covariances, useful for calculating the variances of constructions such as $(\hat{Y}_1^p - \hat{Y}_2^p)$.[2]

The total variability is again merely the sum of the estimation and fundamental variabilities.

An Example

Consider the Poisson regression model (King, 1989: Chapter 5). The stochastic component is Poisson:

$$Y \sim \frac{e^{-\lambda}\lambda^y}{y!} \tag{20}$$

and the systematic component is exponential:

$$\lambda = \exp(X\beta) \tag{21}$$

The likelihood function is then

$$L(\beta|y) = \prod_{i=1}^{n} \frac{\exp(-e^{X\beta})\exp(X\beta)^y}{y!} \tag{22}$$

The fundamental variability in the Poisson model happens to equal the expected value, so that

$$\begin{aligned} V(Y|\lambda) &= \sum_{y=1}^{\infty} y^2 \frac{e^{-\lambda}\lambda^y}{y!} - \left(\sum_{y=1}^{\infty} y \frac{e^{-\lambda}\lambda^y}{y!}\right)^2 \\ &= (\lambda + \lambda^2) - \lambda^2 \\ &= \lambda \end{aligned} \tag{23}$$

[2] An interesting side point can be studied by extending the Taylor series approximation to a third term, $\frac{1}{2}(b-\beta)'g''(\beta)(b-\beta)$, and calculating $E(\hat{Y}^p)$. The expected value of the second term is zero, and the entire approximation now becomes $E(\hat{Y}^p) \approx g(\beta) + \frac{1}{2}A'V(b)A$, where $A'A = g''(\beta)$, the matrix of second derivatives. (A can be calculated by taking the Cholesky decomposition of the second derivative matrix (in Gauss, for example: `A=CHOL(D)`, where `D` is the second derivative.) This calculation shows the details of the familiar result that $g(E(T)) \neq E(g(T))$, for nonlinear functional forms $g(\cdot)$. However, since the term $\frac{1}{2}A'V(b)A$ goes to zero as the sample size increases, $g(\beta)$ is a reasonable approximation to $E(g(b)) \equiv E(\hat{Y}^p)$ in large samples.

Where the summation is used in place of integration since the distribution is discrete. The vector λ is $(N \times 1)$. By the usual no autocorrelation assumptions of the Poisson regression model, the covariances (based on fundamental variability) between different predicted values are zero. Thus, one can write the $(N \times N)$ fundamental variance matrix as λI, where I is an $(N \times N)$ identity matrix. To estimate this fundamental variability, one would use the $(N \times N)$ matrix $\exp(X^p b)I$.

To calculate the estimation variability, we need the first derivative matrix

$$\begin{aligned} g'(\beta) &= \frac{\partial \exp(X\beta)}{\partial \beta} \\ &= X \cdot e^{X\beta} \end{aligned} \tag{24}$$

Note that this is a slight abuse of standard mathematical notation, used in order to make the transition to computation easier. X is $(N \times k)$ and $e^{X\beta}$ is $(N \times 1)$. The notation "$\cdot$" in this equation refers to multiplying each column of X by $e^{X\beta}$ element by element.[3]

Thus, the estimated variance matrix of the k-vector of predicted values $\hat{Y}^p$ is then as follows:

$$V(\hat{Y}^p) = (X^p \cdot e^{X^p b})\widehat{V(b)}(X^p \cdot e^{X^p b})' \tag{25}$$

One would add $\hat{\lambda}$ or, equivalently, $\exp(X^p b)I$ to the diagonal elements of this matrix in order to calculate the total variability.

Concluding Remarks

An important way to evaluate a statistical model is to calculate the total variability and see whether roughly two-thirds of the observed points are within plus or minus a standard error of the predicted values, and whether about 95% of the observations are within plus or minus two standard errors. Without a calculation such as this, all the standard errors and statistical tests based on this model would be in doubt.

Reference

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

Johnston, J. 1984. *Econometric Methods.* 3rd edn. New York: McGraw Hill.

[3] In Gauss, one would use `X.*exp(X*b)`.

Maximum Likelihood for the Masses

Donald Green
Yale University

What is maximum likelihood estimation? A great many textbooks address this question, but few do so in ways that would be accessible to the typical political science graduate student, whose calculus background is rusty or nonexistent.[1] My sense is that students would better understand maximum likelihood (MLE) estimation—and might warm up to the more formal treatments of the topic—if they could see how it works computationally. The examples I discuss below are intended to cover a broad range of applications, while at the same time requiring a minimal knowledge of calculus or advanced statistics. Machine-readable copies of the GAUSS programs presented here are available from the author (BITNET address: gogreenyalevm).

A Simple Introductory Example

If we flip a coin six times (and each flip is independent of the others), what are the odds of obtaining two heads? The answer depends on whether the coin is biased or not. Recall from introductory statistics (Wonnacott and Wonnacott, 1984: 94-5) that if the coin is fair ($p[heads] = .5$), the likelihood (L), or probability, of obtaining $k = 2$ heads in $n = 6$ flips is

$$\begin{aligned} L &= \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \\ &= .234375. \end{aligned}$$

That is, when we flip a fair coin six times, we have roughly a 1 in 4 chance of coming up with exactly two heads. But what if the coin were biased and the probability of heads were .2? Then the likelihood of obtaining 2 heads increases to .24576. Trying out some different values leads us to conclude that a probability of heads equal to $\frac{1}{3}$ yields the largest likelihood of all: .329218. In other words, a biased coin with $p[heads] = \frac{1}{3}$ would maximize the likelihood of observing 2 heads in 6 tries.

Note a subtlety here. We're not saying that the coin is likely to have a $p[heads] = \frac{1}{3}$ or even that the most likely probability of obtaining heads with this coin is $\frac{1}{3}$. The likelihood properly refers to the observed outcome (2 heads out of 6), not the coin. To say that $\frac{1}{3}$ maximizes the likelihood of obtaining the data is simply to say that the odds of obtaining the observed result are greatest when $p[heads] = \frac{1}{3}$. At first glance, this aspect of maximum likelihood seems unsatisfying; after all, we generally want to form some assessment about the coin itself. The point to remember, however, is that maximum likelihood is simply an estimation procedure—one of many potential ways in which one might generate a guess as to what the true $p[heads]$ is. It turns out that maximum likelihood estimates have many desirable properties, especially when the sample size is fairly large (King, 1989).

Think of L as a function – that is, a quantity whose value depends on the number of flips (n), heads obtained (k), and p, the true probability of obtaining heads with that coin. When n and k are given, as in the foregoing example, it's not hard to find the value of p—call it p_{ML}—that maximizes this function simply by trying out different potential values of p.[2] But when the number of parameters is greater than one, this "grid search" technique becomes tedious. Since we're interested in figuring out how to solve maximization problems in general, it pays to invest some time learning about the OPTMUM routines in GAUSS.[3]

When using programs such as OPTMUM, it is conventional to maximize the natural logarithm of L rather than L itself. The reason is that the expression for Log-L is easier to manipulate algebraically and computationally. Conceptually, replacing L with Log-L has no significance, since the maximum value of L will also maximize Log-L. But now the function to be maximized is a bit more tractable:

$$ln(L) = ln\left(\frac{n!}{k!(n-k)!}\right) + (k)ln(p) + (n-k)ln(1-p)$$

The GAUSS code listed in Listing 1 shows how OPTMUM can be used to maximize this function. The program begins by calling in the programs contained in the OPTMUM library. Next, values for the "data" are inserted, $n = 6$ and $k = 2$. A starting value is provided (called $p0$), so that OPTMUM knows where to begin its iterative search for the maximum of the log-likelihood function. A procedure, called "qfct(p)," is designed to return the value of log-L for various guesses of p. Finally, the OPTMUM command

[1] For a basic introduction, see Wonnacott and Wonnacott (1984: 499-514). Johnston (1984: 274-9) provides a more technical discussion of maximum likelihood with reference to linear regression. Judge et al. (1988) cover nonlinear models, simultaneous equations, censoring and truncation. Bollen (1989: 136-144) illustrates how maximum likelihood may be used to generate LISREL estimates. Rothchild and Logothetis (1986) and King (1989) provide a useful overview of different maximum likelihood estimators.

[2] In this case, the search process can be speeded considerably by taking the derivative of L with respect to p, setting the result equal to zero, and solving for p. This procedure yields a closed-form expression for the maximum likelihood estimate of p, namely $\frac{k}{n}$. In many cases, however, no such closed-form solution will be forthcoming.

[3] Many statistical packages have routines specifically designed to perform maximum likelihood estimation. GAUSS is no exception, and readers may find it more convenient to use MAXLIK rather than OPTMUM. We have opted to use OPTMUM because we wish to make transparent some of the routines that are done automatically in MAXLIK. For GAUSS code that is even more stripped-down than what we provide here, see the maximum likelihood programs presented in Hill (1989, 112-141).

is called (see Chp. 12 in the GAUSS Applications Manual), standard errors calculated, and estimates printed.

When this program is executed, we obtain three pieces of output: a p_{ML} of .33333, a standard error for p_{ML} of .19245, and a log-likelihood of -1.11103. The first number is what we obtained above. The log-L value is simply the log of .329218, which, as we noted above, is the value of the likelihood function evaluated at p_{ML}=.33333. The only new number is the standard error, which provides a rough sense of the sampling variability of the estimate. I say "rough sense" because the standard errors are only approximations in small samples.

Using Maximum Likelihood to Perform Linear and Nonlinear Regression

The beauty of maximum likelihood estimation is its flexibility. With only a few modifications to the program just presented, we can get OPTMUM to do linear and nonlinear regression analysis. Consider, for purposes of illustration, the basic bivariate regression model:

$$y_i = a + bx_i + u_i$$

where x_i is the independent variable, y_i is the dependent variable, and u_i is the error term. The parameters a and b, the intercept and slope, are the constants we wish to estimate using maximum likelihood. This is relatively easy to do as long as we're prepared to make some assumption about the distribution of u_i. If we assume that the u_i are drawn from a normal population with mean 0 and variance σ^2, then the log-likelihood function for the data is the following (see Johnston 1984, p.271):

$$ln(L) = -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma^2) - \left(\frac{1}{2\sigma^2}\right)\sum_{i=1}^{n} c_i^2$$

where

$$c_i = y_i - a - bx_i.$$

The guesses for a and b that maximize this function are the maximum likelihood regression estimates. Listing 2 presents an OPTMUM program to run MLE on some simulated data. An OLS analysis is also run for purposes of comparison. The regression estimates are identical, but the standard errors are different. The fault lies with MLE; its standard errors are biased in small samples.

The extension to nonlinear regression is a lot easier than one might think. Suppose we wanted to estimate the degree of curvature in the relationship between x_i and y_i, as well as the slope and intercept. A Box-Cox model (Johnston 1984, p.62) would look like the following:

$$y_i = a + b\left(\frac{x_i^\lambda - 1}{\lambda}\right)$$

Note that as $\lambda \rightarrow 1$, y_i becomes a linear function of x_i; as $\lambda \rightarrow 0$, y_i becomes a linear function of $ln(x_i)$. The log-likelihood function is the same as (4), although (5) is now

$$e_i = y_i - a - b\left(\frac{x_i^\lambda - 1}{\lambda}\right)$$

Listing 3 presents an OPTMUM program to perform a nonlinear regression on 500 simulated observations. Because the true model used to generate the data is in fact linear in the x_i, the true $\lambda = 1$. The estimate λ_{ML}, .98671, comes quite close. For purposes of comparison, I also included an OLS regression that assumes $\lambda = 1$.

A word about starting values. The values I supplied for "$p0$" were "good" in the sense that OPTMUM was able to converge to a final set of estimates. You may find that maximum likelihood routines at times go belly-up before converging upon a final set of estimates. In other words, in their iterative search for the parameter values that maximize the likelihood function, routines sometimes try out values that are for one reason or another inadmissible (e.g., values that make the likelihood function negative, hence the log-likelihood becomes undefined), and the program crashes. This problem is annoying but not insurmountable. A useful technique is to call one of the parameters that would otherwise be estimated a fixed constant (say z). Set z to some plausible value and choose starting values for the remaining "free" parameters. Run this "constrained" MLE program for different values of z and record the value of Log-L after each run. (The program may still crash because your other starting values aren't good enough; keep guessing until the program gives you a break, then change z and the other starting values very gradually.) When you seem to find the value of z that maximizes Log-L, use the parameter estimates from this run as starting values. For example, you might run the program in Listing 3, replacing $p[3,1]$ with z, which is globally set to .90 and repeat this procedure for various values of z. When you get the sense that your three starting values are "good", try them out in the unconstrained MLE program. Starting values can be a headache, but it's really just a matter of perseverance.

The Public's Willingness to Pay for Artistic and Cultural Activities

Public officials are continually impelled to extend the range of government services but, especially in an era of fiscal retrenchment, are hard pressed to come up with the funds to pay for new programs. One obvious solution is to raise direct and indirect taxes. This is not without political risk, of course, particularly when the levy exceeds what most of the citizenry is willing to pay. For this reason, public officials and advocacy groups take interest in polls that ask questions, in one form or another, about the public's willingness to pay more in taxes in order to expand government

programs (see Green 1991). We consider for purposes of illustration a question posed to a national sample of adults in 1973 concerning the arts:

> Would you be willing to pay an additional $50 a year in your taxes if the money were used to maintain and operate cultural facilities such as theatre, music, and art exhibitions? (if would not or not sure, ask:) How about a tax rise of $25 to pay for such facilities, would you be willing to pay that much, or not? (if would not or not sure, ask:) Would you be willing to pay an additional $5 in taxes to make such activities more available?

Let w_i be a random variable representing the maximum amount that a person i is willing to pay to support the arts. Let $\{p_1, p_2, p_3\}$ represent the three prices offered, and let $\{I_1, I_2, I_3, I_4\}$ represent the number of respondents that fall in the various intervals between and outside the 3 price thresholds, so that I_1 people refuse to pay p_1, I_2 refuse to pay p_2 but agree to pay p_1, etc. The likelihood function for these data is

$$\begin{aligned} L &= Pr\left[w < p_1\right]^{I_1} + \\ & \quad Pr\left[p_1 \leq w < p_2\right]^{I_2} + \\ & \quad Pr\left[p_2 \leq w < p_3\right]^{I_3} + \\ & \quad Pr\left[p_3 \leq w\right]^{I_4} \end{aligned}$$

The log-likelihood function is therefore

$$\begin{aligned} lnL &= I_1 lnPr\left[w < p_1\right] + \\ & \quad I_2 lnPr\left[p_1 \leq w < p_2\right] + \\ & \quad I_3 lnPr\left[p_2 \leq w < p_3\right] + \\ & \quad I_4 lnPr\left[p_3 \leq w\right] \end{aligned}$$

Before we can maximize this function, we need to provide more information about the probability distribution from which the data are drawn. In the coin-flipping example, it was clear that the data were generated from a binomial distribution. In more complex applications, deciding on an underlying distribution (e.g., normal, uniform, Poisson, etc.) tends to be the most difficult and contentious aspect of maximum likelihood. In the present example, we have the good fortune to be able to rely on extensive scholarship on the public's willingness to pay for so-called "public goods." This literature indicates that w_i generally follows a logistic distribution when the asking price is transformed into the natural logarithm of price. Therefore, let $\{p_1, p_2, p_3\}$ = {$5,$25,$50}. Assuming w_i to be distributed logistically implies the following likelihood function:

$$\begin{aligned} lnL &= I_1 ln\left[\frac{e^{-a-bp_1}}{1+e^{-a-bp_1}}\right] + \\ & \quad I_2 ln\left[\frac{e^{-a-bp_2}}{1+e^{-a-bp_2}} - \frac{e^{-a-bp_1}}{1+e^{-a-bp_1}}\right] + \\ & \quad I_3 ln\left[\frac{e^{-a-bp_3}}{1+e^{-a-bp_3}} - \frac{e^{-a-bp_2}}{1+e^{-a-bp_2}}\right] + \\ & \quad I_4 ln\left[\frac{1}{1+e^{-a-bp_3}}\right] \end{aligned}$$

As before, maximum likelihood estimates are generated by selecting values of a and b that maximize this likelihood function.

The GAUSS code used to estimate these two parameters, the intercept and effect of price on willingness to pay, is presented in Listing 4. Presumably, the slope (b) is negative, indicating that public support wanes as the asking price increases. The log-likelihood function to be maximized is the same as in eq. 10. The starting values for the iterative parameter search are 1.2 and -.7. The final maximum likelihood estimates are 1.63803 and -.52999. The intercept may be interpreted to mean that 83.7% of the public would have been willing to pay $1 more in taxes to help the arts.[4] The slope implies that for each one-unit change in the log of price, the log-odds of willingness to pay declines by .53.[5] Substantively, this may be interpreted to mean that willingness to pay declines rather slowly as the asking price increases. As price rises by an order of magnitude from $5 to $50, our model predicts a drop in public support from 68.7% to 39.3%.

How close is the statistical fit between hypothesized model and the data? One way to answer this question is to compare the observed and expected frequencies within each of the price intervals. Expected frequencies are calculated by using the maximum likelihood parameter estimates to estimate the probability of drawing observations in each interval, then multiplying by the sample size (see Listing 4). Casual inspection of the output indicates that the two outer intervals are well predicted, while the inner intervals are off. This intuition is confirmed by a χ^2 test of fit. The value of the χ^2 statistic for this model with 1 degree of freedom (3 independent pieces of information minus 2 parameters to be estimated) is 18.8; the prob value is .000.

Why the poor fit? In one sense the fit is not all that bad: 18.3% of the cases fell into the second interval; 20.4% were expected. The margin of error in other words is not very great, but since the sample is large ($n = 2785$) minor deviations from the hypothesized distribution produce a low prob value. One could switch to a different distribution (e.g., lognormal or Weibull) in the hopes of reducing the χ^2. Alternatively, one could relax the assumption that price follows a natural log function by introducing a λ exponent

[4] Since $Pr\left[\$1 \leq w\right] = \frac{1}{1+e^{[-a+bln(1)]}} = \frac{1}{1+e^{[-1.63803]}} = 83.7$.

[5] Log-odds are not as abstruse as they first appear. The log-odds of 75% willingness to pay for a given price, for example, implies an odds of 3:1 and a log-odds of ln(3/1). In order to translate log-odds (denoted q) into percentages (denoted v), use the formula $v = \frac{1}{1+e^{[-q]}}$.

in the spirit of a Box-Cox model (see above).[6]

To this point we have avoided discussion of how the standard errors are generated because the subject is a bit technical. (Readers averse to technical stuff should skip to the end of the paragraph.) The GAUSS procedure HESSP in this example calculates a 2x2 matrix of partial second derivatives of the log-likelihood function with respect to the parameters, evaluated at the maximum likelihood estimates of these parameters. The negative inverse of this matrix, called "invmat" in Listing 4, contains the variances and covariances of the parameter estimates. Thus, for example, the standard error for the second parameter estimate, $\hat{b}$, may be found by taking the square root of the [2,2] element in the "invmat" matrix. An important word of caution is in order, however. The fine print of the GAUSS Command Reference (p.1218) warns that the numerical second derivatives can be flaky. It's always a good idea to print out the gradients (a 2x1 vector called "g" in my program). If the elements differ by more than a factor of "100 or so," beware. In the present example, the gradients differed by only a factor of 8.3.

Conclusion

The numerical maximization of likelihood functions is relatively straightforward once one has a working knowledge of OPTMUM. Less easy is the algebraic derivation of likelihood functions from a set of distributional assumptions. But many widely used derivations are readily available from textbooks (see note 1). Furthermore, the advent of computer programs such as Mathematica, which solve calculus problems easily and quickly, make technical problems of this kind increasingly trivial. The real challenge to those wishing to use maximum likelihood is ultimately neither computational nor algebraic. Rather the most difficult aspect of using MLE is to come up with distributional assumptions that are appropriate for a given substantive application. For guidance here, readers may wish to consult King (1989) and Rothchild and Logothetis (1986).

References

Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York: John Wiley & Sons.

Green, Donald Philip. The Price Elasticity of Mass Preferences. Yale Institution for Social and Policy Studies Working Paper No. 1041.

Johnston, J. 1984. *Econometric Methods*, 3rd Ed. New York: McGraw-Hill.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, and Tsoung-Chao Lee. 1988. *Introduction to the Theory and Practice of Econometrics*, 2nd Ed. New York: John Wiley & Sons.

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

Rothchild, V. and N. Logothetis. 1986. *Probability Distributions*. New York: John Wiley & Sons.

Wonnacott, Thomas H. and Ronald J. Wonnacott. 1984. *Introductory Statistics for Business and Economics*, 3rd Ed. New York: John Wiley & Sons.

[6] That is, instead of using the log of price, one might use the Box-Cox transformation presented in eq. 6. Doing so yields a perfect fit to the data since the model is just identified (there are three independent price thresholds and three parameters to be estimated). The value of λ_{ML} is .272 (SE=.064), indicating that the relationship between the log-odds of willingness to pay and price is not as curvilinear as the log transformation implies. For additional evidence on this point, see Green (1991).

Listing 1: Estimation of a Binomial Parameter

```
      @        Maximum Likelihood:       @
      @ Estimating a binomial parameter  @
      @ --------------------------------@
library optmum;  @ calls in compiled code @
                 @ from GAUSS-386 @

n=6;                            @ # of trials    @
k=2;                            @ # of successes @

let p0= {.35};   @ provides  starting value;   @
    @ for best results, this should be "close" @

@ what follows is the log-likelihood f(x) @

proc  qfct(p);
 retp( ln(n!/(k!*(n-k)!)) +
       ln(p)*k + ln(1-p)*(n-k)  );
endp;

@Call OPTMUM and return parameter estimates@
{p,f,g,h}=optmum(p0,&qfct);

output file="binom.out" reset; format 10,5;

@ take 2nd partial derivative of likelihood    @
@ function w/ respect to p, evaluated at p-hat @
hessmat=hessp(&qfct,p);

@ take the negative inverse of the hessian      @
@ and print out the square root of the diagonal @
invmat=inv(-1*hessmat);

"log-likelihood function at maximum=";;f; print;
print p' "= Maximum likelihood parameter estimate";
print sqrt(invmat)
"= Standard errors of the ML estimate";
output off;

***********************************************
      Output for Binomial Estimation Run:
```

```
************************************************

log-likelihood function at maximum=  -1.11103

0.33333 = Maximum likelihood parameter estimate
0.19245 = Standard errors of the ML estimate
************************************************
```

Listing 2: Estimating a Normal Linear Model

```
      @       Maximum Likelihood:         @
      @  Estimating a normal linear model @
      @ --------------------------------- @
library optmum;
        @ set up the simulation @
cases=20;          @ the number of cases (n)     @
seed=573424264;    @ sets seed for random numbers@
x=sqrt(10)* rndus(cases,1, seed); @ generates x  @
            @ generates normal disturbances (u) @
u=sqrt(5) * rndns(cases,1, seed);

c=ones(cases,1);       @ generates the intercept @
     @ true law for y: a=5, b=-2, sigma-sqared=5 @
y=5*c + -2*x + u;

     @ starting values for a, b, and s-squared   @
let p0= {5,-2,5};


@ the log-likelihood function for the data  @
@ note that p[1,1] represents a             @
@ note that p[2,1] represents b             @
@ note that p[3,1] represents sigma-squared @

proc  qfct(p);

   retp(-(cases/2)*ln(2*pi) -
   (cases/2)*ln( p[3,1] )-(1/ (2* p[3,1]) )
   * (y-(c~x)*(p[1,1] | p[2,1]))'(y-(c~x)
   * (p[1,1] | p[2,1])));

endp;  @ End of Likelihood function @

@ Call OPTMUM and Return Parameter Estimates @
{p,f,g,h}=optmum(p0,&qfct);

output file= "ols.out" reset; format 10,5;

     @ calculate 3 x 3 matrix of 2nd     @
     @ partial derivatives of the log-   @
     @ likelihood function w/ respect to @
     @ the parameters, evaluated at the  @
     @ parameter estimates               @
hessmat=hessp(&qfct,p);

     @ take the negative inverse of the  @
     @ hessian                           @
invmat=inv(-1*hessmat);

     @ and calculate the standard errors @
std=sqrt(diag(invmat));

     @ Print the results @
"log-likelihood function at maximum =" f;
print;
print " Parameter   Standard";
print " Estimates     Errors";
p~std;
print;
print "------------ OLS results -------------";
ols(0,y,x); @ Run OLS on data @
output off;
```

```
************************************************
         Output for Normal Linear Model
************************************************
log-likelihood function at maximum = -46.29327

  Parameter   Standard
  Estimates    Errors

   5.28849    1.22601
  -2.03732    0.64207
   5.99813    1.89681

 ------------ OLS results -------------
                           Standard
Variable     Estimate       Error
-----------------------------------
CONSTANT     5.288492    1.292332
X1          -2.037319    0.676806
```

Listing 3: Estimating a Normal Nonlinear Model

```
  @           Maximum Likelihood:             @
  @ Estimating a normal nonlinear model       @
  @ ----------------------------------------- @

library optmum;
cases=500;               @ number of cases (n) @
seed=57344264;    @ sets seed for random numbers @
x=sqrt(10)* rndus(cases,1, seed); @ generates x  @
     @ generates normal disturbances (u)  @
u=sqrt(5) * rndns(cases,1, seed);
c=ones(cases,1);       @ generates the intercept @
@ true law for y: a=5, b=-2, sigma-sq=5,lambda=1 @
y=5*c + -2*x + u;
   @ starting values for a, b, s-sq, lambda @
let p0= {5,-2,5,1};

  @ The log-likelihood function for the data  @
  @ p[1,1] represents a                       @
  @ p[2,1] represents b                       @
  @ p[3,1] represents sigmat-squared          @
  @ p[4,1] represents lambda                  @
proc  qfct(p);
```

```
 retp(-(cases/2)*ln(2*pi)
      -(cases/2)*ln( p[3,1] ) -(1/ (2* p[3,1]))
  * (y-p[1,1] - p[2,1]* (x.^p[4,1] - 1)./p[4,1])'
  (y-p[1,1] - p[2,1]* (x.^p[4,1] - 1)./p[4,1]));

endp;

@ Call OPTMUM and Return Parameter Estimates @
{p,f,g,h}=optmum(p0,&qfct);

output file= "box.out" reset; format 10,5;

     @ calculate 4 x 4 matrix of 2nd     @
     @ partial derivatives of the log-   @
     @ likelihood function w/ respect to @
     @ the parameters, evaluated at the  @
     @ parameter estimates               @
hessmat=hessp(&qfct,p);

     @ take the negative inverse of the  @
     @ hessian                           @
invmat=inv(-1*hessmat);

     @ and calculate the standard errors @
std=sqrt(diag(invmat));

     @ Print the results @
"log-likelihood function at maximum =" f;
print;
print "  Parameter   Standard";
print "  Estimates    Errors ";
p~std;
print;
print " ------------ OLS results -------------";

     @ Run an OLS regression: equivalent  @
     @ to a nonlinear regression assuming @
     @ p[4,1] = 1.00                      @
call ols(0,y,x);
output off;

*************************************************
       Output for Normal Nonlinear Model
*************************************************
log-likelihood function at maximum =-1117.26663

  Parameter   Standard
  Estimates    Errors

   2.97399    0.14566
  -1.83281    0.14203
   5.10995    0.32319
   0.98671    0.16880

------------ OLS results -------------
                         Standard
Variable     Estimate       Error
-----------------------------------
CONSTANT     4.806666    0.198752
X1          -1.825876    0.112141
```

Listing 4: Estimation of a Logistic Choice Model

```
   @           Maximum Likelihood:               @
   @ Application to Estimation of a Logistic     @
   @ Willingness to pay for Arts &               @
   @ Cultural Activities                         @
   @ Branched Question format                    @
   @ Data are from Green (1991)                  @
   @ ------------------------------------------- @
library optmum;
i1=885 ;  @ # willing to pay less than $5        @
i2=511 ;  @ # willing to pay between $5 and $25  @
i3=308;   @ # willing to pay between $25 and $50 @
i4=1081;  @ # willing to pay at least $50        @

@ starting values for iterative parameter search @
let p0= {1.2,-.7};

@ transform asking prices by taking natural logs @
 price1=ln(5);
 price2=ln(25);
 price3=ln(50);

   @ log-likelihood function to be maximized     @
   @ p[1,1] represents the intercept             @
   @ p[2,1] represents the slope                 @

proc  qfct(p);
  retp(ln( (    exp(-p[1,1] - price1*p[2,1]) ) /
      ( 1+ exp(-p[1,1] - price1*p[2,1]) ) ) * i1 +
  ln( ( (  exp(-p[1,1] - price2*p[2,1]) ) /
      ( 1+ exp(-p[1,1] - price2*p[2,1]) ) ) -
    ( (    exp(-p[1,1] - price1*p[2,1]) ) /
      ( 1+ exp(-p[1,1] - price1*p[2,1]) ) ) )*i2 +
  ln( ( (  exp(-p[1,1] - price3*p[2,1]) ) /
      ( 1+ exp(-p[1,1] - price3*p[2,1]) ) ) -
    ( (    exp(-p[1,1] - price2*p[2,1]) ) /
      ( 1+ exp(-p[1,1] - price2*p[2,1]) ) ) )*i3 +
  ln(   1                              /
      ( 1+ exp(-p[1,1] - price3*p[2,1]) ) )   * i4
               );
endp;


@ Call OPTMUM and Return Parameter Estimates @
{p,f,g,h}=optmum(p0,&qfct);
output file="art73.out" reset;format 10,5;

   @ calculate 2 x 2 matrix of 2nd    @
   @ partial derivatives of the Log-  @
   @ likelihood function with respect @
   @ to the two parameters            @
hessmat=hessp(&qfct,p);
```

```
   @ take the negative inverse of the @
   @ hessian matrix                   @
invmat=inv(-1*hessmat);

   @ calculate standard errors        @
std=sqrt(diag(invmat));

     @ Print the results @
print "log-likelihood function at maximum =" f;
print;
print "parameter    standard ";
print "estimates     errors  ";
p~std;
print;

@ Perform a chi-square test for goodness of fit @
   @ compute the total n in the sample   @
   @ and the expected n in each interval @
n=i1+i2+i3+i4;

a1hat=(  ( exp(-p[1,1] - price1*p[2,1]) ) /
       ( 1+exp(-p[1,1] -price1*p[2,1]) ) )    * n;
a2hat=( ( (exp(-p[1,1] - price2*p[2,1]) ) /
      ( 1+ exp(-p[1,1] - price2*p[2,1]) ) ) -
    ( (    exp(-p[1,1] - price1*p[2,1]) ) /
      ( 1+ exp(-p[1,1] - price1*p[2,1]) ) ) ) * n;
a3hat=( ( (exp(-p[1,1] - price3*p[2,1]) ) /
      ( 1+ exp(-p[1,1] - price3*p[2,1]) ) ) -
    ( (    exp(-p[1,1] - price2*p[2,1]) ) /
      ( 1+ exp(-p[1,1] - price2*p[2,1]) ) ) ) * n;
a4hat=( 1                              /
      ( 1+ exp(-p[1,1] - price3*p[2,1]) ) )   * n;
print;
print "Observed vs. Expected in each interval";
print;
print i1 a1hat " |          WTP < $5" ;
print i2 a2hat " |    $5 <= WTP < $25";
print i3 a3hat " |   $25 <= WTP < $50";
print i4 a4hat " |          WTP > $50";

   @ calculate chi-square @
chi=(i1-a1hat)^2/a1hat+(i2-a2hat)^2/a2hat+
    (i3-a3hat)^2/a3hat+(i4-a4hat)^2/a4hat;
print;
chi " chi-square with 2 d.f.";
cdfchic(chi,2) " prob-value";
output off;

************************************************
         Output for Log-Logistic Model
************************************************

log-likelihood function at maximum =-3591.33269

parameter    standard
estimates     errors

   1.63803    0.05933
  -0.52999    0.01654

Observed vs. Expected in each interval

 885.00000  872.35797  |          WTP < $5
 511.00000  567.43214  |    $5 <= WTP < $25
 308.00000  251.09695  |   $25 <= WTP < $50
1081.00000 1094.11294  |          WTP > $50

  18.84789  chi-square with 2 d.f.
   0.00008  prob-value
```

Event Count Models for Supreme Court Dissents

David C. Nixon
Washington University

Introduction

Ordinary Least Squares (OLS) is probably the most often used statistical procedure in political science. But some topics are more appropriately analyzed through other procedures. Much of the data analyzed in political science are discrete measures of event counts. We have strong theoretical reasons for expecting event counts to have a poisson distribution. In addition, we have strong theoretical reasons for distinguishing between normal and poisson distributions in small samples.

In spite of these reasons, most researchers immediately run an OLS regression on their data and seldom go on to more sophisticated methods. But OLS is not well suited to event count data because event counts are discrete and do not take on negative values, while OLS predicted values are continuous and may take on negative values. Solutions to this problem have been discussed by King (King 1988, 1989, 1989b). King also presents an account of the logged-OLS (LOLS) procedure. He points out some of the flaws of this procedure and presents two maximum likelihood estimation (MLE) solutions to the problems of event count data. Since MLE forms the core for this issue of *The Political Methodologist*, I thought it would be interesting to look at a substantive application of event count models and the pitfalls possible in such procedures.

Estimation Models

The simplest and most common estimation procedure involves an OLS regression of the dependent variable, y, on the matrix of independent variables, x. However, linear OLS models may predict negative numbers for y and this outcome is nonsensical for event counts. One tactic has been to fit the curve in the area of the data. In such cases, we ignore the negative predictions or truncate them at zero.

This is clearly an undesirable and ad hoc solution. Another tactic has been to substitute an exponential function:

$$y = exp(x\beta + u) \quad (1)$$

where u is a random error term and $x\beta$ represents the matrix of independent variables and their coefficients. Equation 1 is equivalent to:

$$\ln(y) = x\beta + u \quad (2)$$

This tactic provides two advantages. First, it eliminates the possibility of negative predicted values for y. Second, it creates a nonlinear function for the dependent variable, which seems desirable in the case of event count data. It seems likely that the increase from two to three events has more import than the increase from 39 to 40 events.

However, this form creates significant problems where y takes on values of zero, since ln(0) is undefined. The remedy often taken is to add a small arbitrary constant to y, such as .5, which results in a defined value for $\ln(y)$. The values for $\ln(y+k)$ are presented in table 1 for different values of k.

As King correctly points out, the convention of adding an arbitrary constant is unjustified. Estimates of coefficients and significance tests are substantially affected by the value of the constant whenever y takes on values of zero in the data, so that coefficients are biased. In addition, LOLS still assumes a continuous distribution of y.

As an alternative to LOLS, King suggested the use of MLE where the ys are assumed to be generated from the poisson distribution. The poisson distribution accommodates the truncation at zero and the fact that the distribution of y is discrete. Unfortunately, the model makes an assumption that all events are independent (King 1989b). As a result, the standard errors (SE) on the estimators for poisson MLE may be inconsistent if the assumption of independence is violated.

King developed a generalized event count (GEC) model which does not make the assumption of independence of events. He created a measure of dispersion of events, γ, allowing them to be either overdispersed ($0 < \gamma < 1$) or underdispersed ($\gamma > 1$) and to vary within those ranges. GEC is identical to the Poisson where $\gamma = 1$, so that the Poisson MLE is legitimate where we can assume independence between events. Where $\gamma \neq 1$, the poisson estimators are inefficient and the standard errors of the estimators are inconsistent. GEC provides an estimate of the over or underdispersion and both the estimators and the standard errors are consistent.

An Application

All of this suggests our choice of models may have substantive implications and that less appropriate procedures may provide misleading views of the real processes underlying the data. It may be useful to look at an application of various procedures to event count data.

In the aftermath of the swearing in of a new justice to the Supreme Court, it is interesting to examine a hypothesis that has been floating around in judicial circles, reappearing occasionally, for over three decades— the freshman effect hypothesis. The original theory, proposed by Eloise Snyder, suggested that new justices are subject to group influences on the Court and they alter their behavior in some way so as to work their way into the Court machinery while antagonizing as few other justices as possible (Snyder 1958). This hypothesis has resurfaced several times in the past few years, most recently in Allen, 1991. Without getting into the theory too heavily, I hypothesized that new Justices would be less likely to dissent on cases their first year than in later years.

I adopt a relatively simple model of Supreme Court Justice dissent behavior. I assume that dissents are a function of a freshman effect, tenure on the court, ideological agreement with the majority, court workload, and year.

The judicial literature has indicated a general trend towards greater dissents during a Justice's tenure. Such a trend may be due to the inevitable changes in the composition of the Court over time. Consider, for example, a Justice nominated by Hoover in 1932 who later must vote on a court dominated by Roosevelt appointees. Because dissents may not increase throughout a Justice's tenure, I include both tenure and tenure squared, to allow for potential decline in dessents in later years.

The judicial literature has also repeatedly indicated the significance of political party as a determinant of behavior of justices. New members coming on to a court whose majority is constituted by the opposing political party are likely to dissent more often. A dummy variable indicates whether the Justice's own political party constituted the majority of the Court.

It is reasonable to suggest that workload has an effect on dissent behavior. This is also a necessary component to include in event count models where the intervals between observations is not constant (King 1989).

Finally, the argument has been made that dissent rates have increased over the twentieth century as the nature and scope of the Court's work has changed and expanded. To capture this I include year as an independent variable.

A variable measuring the number of events occurring during a fixed time period can be expected to have a poisson distribution:

$$y \sim f_{\text{Poisson}}(\lambda) \quad (3)$$

where λ is the rate of event occurrence. It would be implausible to assume that all justices have the same λ, and indeed, it is variation in dissent rates across justices which is the focus of the substantive concerns of the model. To capture these substantive influences, I assume that λ is a

function of the variables discussed above. Specifically, I assume that

$$\lambda = exp(x\beta)$$

where x is the matrix of independent variables. Using the event count estimation package COUNT, developed by King, we can easily estimate the parameters of the model and their standard errors.

Results

For comparison, I present results for both the LOLS and the event count models. The use of the LOLS model requires a choice of k, the constant added to y to avoid zero values. This may not be an inconsequential choice. For the LOLS model, I find a strikingly large effect of k on the substantive results. The data set contained only seven out of 383 cases where $y = 0$, so one might expect the choice of k to play little role. Table 1 illustrates these effects for the freshman coefficient. As k decreases, the standard errors for the freshman coefficient increase dramatically. While the coefficient varies only modestly, the substantive conclusions are reversed, from support for the freshman effect hypothesis to no support.

Table 1: Logged OLS regression

k	$\ln(k)$	Freshman coefficient	SE	Prob (1 tail)
.5	-0.69	-.327	.178	.038
.1	-2.30	-.342	.211	.053
.01	-4.61	-.344	.263	.096
.001	-6.91	-.341	.321	.145
.0001	-9.21	-.337	.384	.190
.0000001	-16.12	-.328	.586	.289

Perhaps most importantly, we cross a significance threshold for the freshman effect between the two most common values of k, .1 and .5. The choice of k, which is completely arbitrary, has substantive implications using our data where zero values for the dependent variable constituted less than 2% of our cases!

Table 2 presents the results for the full model estimated with LOLS, poisson event count and Generalized Event Count models. (The LOLS uses $k = .5$.)

As expected, we obtain significantly different results, depending on the event count model we choose. The poisson model assumes $\gamma = 1$ and this assumption seriously affects the coefficients.

We find, not surprisingly, that dissent rates on the Supreme Court are significantly overdispersed. That is, a dissent negatively affects the likelihood of another dissent in the near future. This fact interacts with the freshman dummy enough to cross a significance threshold between the GEC and poisson models.

Table 2: Event Count Models for Dissents

	LOLS	Poisson	GEC
Constant	-3.639	-2.109	-1.382
	(1.529)	(.497)	(1.176)
Tenure	.072	.064	.048
	(.021)	(.007)	(.016)
Tenure^2	-.002	-.002	-.0011
	(.001)	(.000)	(.0005)
Freshman	-.327	-.351	-.317
	(.178)	(.077)	(.169)
Majority	-.107	-.030	-.076
	(.086)	(.027)	(.065)
Year	.047	.036	.032
	(.006)	(.002)	(.004)
ln(Decisions)	.755	.592	.497
	(.266)	(.087)	(.207)
γ			1.865
			(.078)

The GEC model provides sufficient evidence to reject the null hypothesis on all but the freshman effect and party ideology variables. Signs on the coefficients are as we expected and dissent rates increase with tenure in a nonlinear fashion. Since the coefficient on tenure squared is negative, we predict a leveling off and an actual decline in the number of dissents over the course of a Justice's tenure. Our estimates predict that dissent rates increase up to slightly less than 22 years on the bench, after which they decline.

Dissent rates significantly increased over the course of the period, and we discussed earlier the theoretical reasons for this expectation. Dissent rates significantly increased with rising numbers of total cases being heard.

We are unable to reject the null hypothesis for our measure of ideological conflict on the bench. This should not be too disappointing nor does it indicate ideological conflict has no effect, in principle, on dissent rates. Rather, the results indicate the shortcomings in a unidimensional dummy variable indicator of ideological conflict.

Finally, we are unable to reject the null for the freshman effect hypothesis. It seems that the lower levels of freshman dissent rates which we can observe in the aggregate are best accounted for by the nonlinear growth in dissent rates over time. Dissent rates are not significantly lower during the first year in comparison to a Justice's behavior throughout his tenure. Rather, low first year dissent rates are part of the rapid growth in dissent rates which occurs during the first decade or so of a Justice's tenure.

Conclusions

Our example demonstrates there are empirical reasons to avoid improper models. Choosing the wrong model may mean drawing the wrong conclusions. But even if the results were robust across models, that is a poor justification for the choice of a model. To be sure, OLS procedures are considerably more convenient and accessible to most researchers. MLE, especially GEC, is occasionally tedious and suffers from a certain fickleness in convergence. But the choice of a model should not depend on the convenience of the statistical procedure. Rather, it should be driven by a theory about the underlying processes generating the data. OLS is inappropriate for event counts and LOLS appears to be very unreliable. Poisson MLE represents an improvement by an order of magnitude, but it too may be inappropriate if the events are not independent. GEC allows us to relax our assumptions about the dispersion of events and presents more reliable results.

References

Allen, David W. 1991. "Voting Blocks and the Freshman Justice on State Supreme Courts." *Western Political Quarterly* x:727- 747.

King, Gary. 1988. "Statistical Models for Political Science Event Counts." *American Journal of Political Science* 32:868-93.

King, Gary. 1989. *Unifying Political Methodology.* Cambridge: Cambridge University Press.

King, Gary. 1989b. "Variance Specification in Event Count Models." *American Journal of Political Science* 33:762-84.

Snyder, Eloise. 1958. "The Supreme Court as a Small Group." *Social Forces* 36:232-38.

The COUNT program is available from Gary King. His address is Department of Government, Harvard University, Littauer Center, Cambridge, Massachusetts 02138; BITnet: gmk@harvunxw; Internet: gmk@wjh12.harvard.edu.

A Maximum Likelihood Syllabus

Political science departments are only begining to develop courses in maximum likelihood methods. Until such courses become common, we must rely on other departments or special programs, such as the ICPSR summer program, to teach these methods to our best graduate students. For the last two years, the ICPSR summer program has offered a four week course on maximum likelihood. The course is team taught, most recently by Charles Franklin, Gary King, Nancy Burns, Neal Beck and Lisa Martin. Carla Valle provided both substantive and computing assistance to the students. Below is the syllabus for last summer's course, which was offered during the first session (July) of the ICPSR summer program. Anyone interested in the course can contact Henry Heitowit at ICPSR for information on future offerings. (Henry Heitowit, ICPSR, Center for Political Studies, PO Box 1248, Ann Arbor, MI 48106)

Likelihood Models and Statistical Inference

ICPSR Training Program
Summer, 1991

Charles Franklin, University of Wisconsin–Madison
Gary King, Harvard University
Nancy Burns, University of Michigan
Nathaniel Beck, U.C. San Diego
Lisa Martin, U.C. San Diego
Carla Valle, Harvard University

This course introduces a unified approach to developing new statistical methods for social and political science research from within the likelihood theory of statistical inference. Students will use this underlying theory of inference to "reinvent" known statistical solutions to common problems with social scientific data and to develop models to address the empirical and theoretical problems they might encounter in their research. The course will also consider other approaches to inference, including Bayesian analysis.

The normal pattern of lectures will be Charles Franklin on Mondays and Tuesdays of each week. Gary King will lecture on July 3 and 5; Nancy Burns on July 10–12; Nathaniel Beck on July 23 and Lisa Martin on July 24–26. Franklin will also lecture on July 17. There will be no class on July 18 and 19 due to the annual political methodology conference. Carla Valle will hold open office hours to assist students with both statistical and technical questions.

Grading will be based on students' written and computer assignments. Computer exercises will be set up to use the

GAUSS computer language. The instructors will provide datasets for the exercises, but students may find additional datasets of more interest to themselves.

All readings are to be done during the week indicated as they are keyed to lecture materials. Some of the readings serve as background for the lectures; others explicitly cover the lecture materials. All materials will be on reserve at the ICPSR library located in the Helen Newberry House. All articles and portions of books are available as a mimeo package for the course. In addition, the following book will be available for purchase:

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

Syllabus

Tuesday, July 2—Methods of Estimation—Franklin

- Kmenta, *Elements of Econometrics*, ch. 6

  Discussion of Computer Exercise

Wednesday, July 3—The likelihood theory of statistical inference vs. the Bayesian model of inference—King

- King, chap 1 & 2

Thursday, July 4—No class—enjoy the picnic

Friday, July 5 The probability model of uncertainty and the likelihood model of inference—King

- King, chaps 3 & 4

  Computer exercise comparing OLS to MLE estimation

Monday, July 8—Models for binary dependent variables—Franklin

- King, chap 5, sections 1–3

Tuesday, July 9—Models for ordinal dependent variables Franklin

- King, chap 5, sections 4–6
- McKelvey and Zavoina, "A statistical model for the analysis of ordinal level dependent variables", *J. Mathematical Sociology*, 4 (1975) pp. 103–120.

  Computer exercise estimating probit and logit models.

Wednesday, July 10—Models of tabular data Burns

- King, chap 6.

  Computer assignment on contingency tables

Thursday, July 11—Spatial models I—Spatial heterogeneity, heteroskedasticity, introduction to spatial dependence—Burns

- King, chap 7
- Anselin, *Spatial Econometrics*, chap 9.

  Written assignment on spatial models

Friday, July 12—Spatial models II—models of spatial dependence—Burns

- Anselin, *Spatial Econometrics* chaps 3 & 6
- Doreian, "Estimating Linear Models with Spatially Distributed Data", *Sociological Mcthodology*, 1981, pp. 359–388.

  Computer assignment on spatial dependence

Monday, July 15—Models with nonrandom selection I—Franklin

- King, chap 9, sections 1–3

Tuesday, July 16—Models with nonrandom selection II Franklin

- Amemiya, *Advanced Econometrics*, chap 10.

  Computer assignment on nonrandom selection models

Wednesday, July 17—Statistical inference with missing data—Franklin

- Little and Rubin, *Statistical analysis with missing data*, chaps 1, 5 & 12

  Computer assignment on missing data imputation

Thursday, July 18—No class, methodology meetings

Friday, July 19 No class, methodology meetings

Monday, July 22—Model specification tests—Franklin

- Leamer, "Let's take the con out of econometrics", *American Economic Review*, 73 (1983) pp. 31–43.

Tuesday, July 23 Time series models Beck

- King, chapter 7
- Beck, "Estimating Dynamic Models Using Kalman Filtering", in *Political Analysis*, 1 (1989) pp. 121–157.

Wednesday, July 24—Event count models—Martin

- King, chapter 5, sections 7–10

Thursday, July 25—Expanded event count models Martin

- King, chapter 9, sections 4–6

Computer assignment on event count models

Friday, July 26—Data and models for international relations—Martin

- King, Chapter 8, Section 3
- King, "Event Count Models for International Relations: Generalizations and Applications," *International Studies Quarterly* 33, no. 2 (June 1989): 124–47
- Hufbauer, Schott, and Elliott, *Economic Sanctions Reconsidered: History and Current Policy,* Second Edition (Washington, DC: Institute for International Economics, 1990,), Chapter 5

The Gauss Corner: OPTMUM and MAXLIK Return Codes

Ron Schoenberg
Aptech Systems

The latest versions of MAXLIK and OPTMUM, shipping September 1990, return return codes. Return code 10, is however, undocumented. Here is a short discussion of what this code means, and what to do.

The bread and butter algorithms in both MAXLIK and OPTMUM are the quasi-Newton methods, BFGS and DFP. They are called quasi-Newton because while they compute a Newton-type direction, i.e., the inverse of the Hessian inverse times the gradient, the inverse of the Hessian is actually an updated approximation which does not require inversion. The particular kind of updating method used in BFGS and DFP is called the secant update method. Let H(k) be the approximation of the inverse of the Hessian on the previous iteration. Then the DFP update is

$$H(k+1) = H(k) + pp'/(p'q) - H(k)qq'H(k)/(q'H(k)q)$$

where $p = x(k+1) - x(k)$, $q = g(k+1) - g(k)$, and $x(k)$ is the parameter vector on the k-th iteration, and $g(k)$ is the gradient vector on the k-th iteration. The BFGS update is

$$\begin{aligned} H(k+1) \quad = \quad & H(k) + ((1 + q'H(k)q)/(q'p)) \\ & *(pp')/(p'q) \quad (pq'H \mid Hqp')/(q'p) \end{aligned}$$

The BFGS is closely related to the DFP and is generally regarded to be superior to the DFP.

In the formulas the cross-product $q'p$ occurs in the several denominators. q and p are each differences which means that if the differences approach zero then the update will approach a state of indefiniteness. This state, ironically, can be approached when the minimum is near because in that case the difference in the coefficients and gradients in each step will become very small. MAXLIK and OPTMUM check to see if the differences in p and q get less than 1e-12 and if so it halts the optimization and sets the return code to 10.

What to do? Provided that the relative gradient is reasonably small, at least 1e-3, say, one could regard the function to essentially be at the minimum, or one could use the coefficients at that point as starting values in another run in which the algorithm is set to the NEWTON method which explicitly computes the Hessian.

Examination Questions

Larry Bartels
Princeton University

Analyze the causal theories underlying each of the following arguments. Propose a mathematical formulation that captures the significant aspects of each theoretical statement; do not be concerned about the practical details of your formulation, such as whether the variables are measurable or coefficients estimable.

1. As long as the reason of man continues fallible, and he is at liberty to exercise it, different opinions will be formed. As long as the connection subsists between his reason and his self-love, his opinions and his passions will have a reciprocal influence on each other; and the former will be objects to which the latter will attach themselves. The diversity in the faculties of men, from which the rights of property originate, is not less an insuperable obstacle to a uniformity of interests. The protection of these faculties is the first object of government. From the protection of different and unequal faculties of acquiring property, the possession of different degrees and kinds of property immediately results; and from the influence of these on the sentiments and views of the respective proprietors, ensues a division of the society into different interests and parties.
—James Madison

2. Social condition is commonly the result of circumstances, sometimes of laws, oftener still of these two causes united; but when once established, it may justly be considered as itself the source of almost all the laws, the usages, and the ideas which regulate the conduct of nations; whatever it does not produce, it modifies. If we would become acquainted with the legislation and the manners of a nation, therefore, we must begin by the study of its social condition. —Alexis deTocqueville

3. The totality of relations of production constitutes the economic structure of society, the real foundation, on which arises a legal and political superstructure and to which correspond definite forms of social consciousness. The mode of production of material life conditions the general process of social, political and intellectual life. It is not the consciousness of men that determines their existence, but their social existence that determines their consciousness. At a certain stage of development, the material productive forces of society come into conflict with the existing relations of production or– this merely expresses the same thing in legal terms– with the property relations within the framework of which they have operated hitherto. From forms of development of the productive forces these relations turn into their fetters. Then begins an era of social revolution. The changes in the economic foundation lead sooner or later to the transformation of the whole immense superstructure. —Karl Marx

4. Although the modern hospital is not as a rule operated for profit, it is nonetheless the product of capitalism not only because the capitalist process supplies the means and the will, but much more fundamentally because the capitalist rationality supplied the habits of mind that evolved the methods used in these hospitals. And the victories, not yet completely won but in the offing, over cancer, syphilis and tuberculosis will be as much capitalist achievements as motorcars or pipe lines or Bessemer steel have been. In the case of medicine, there is a capitalist progression behind the methods, capitalist both because to a large extent it works in a business spirit and because it is an emulsion of the industrial and commercial bourgeoisie. —Joseph Schumpeter

5. Take the ideas of the masses (scattered and unsystematic ideas) and concentrate them (through study turn them into concentrated and systematic ideas), then go to the masses and propagate and explain these ideas until the masses embrace them as their own, hold fast to them and translate them into action, and test the correctness of these ideas in such action. Then once again concentrate ideas from the masses and once again go to the masses so that the ideas are persevered in and carried through. And so on, over and over again in an endless spiral, with the ideas becoming more correct, more vital and richer each time. —Mao Tse-tung

Call for Papers: Ninth Political Methodology Conference

Henry E. Brady
University of California, Berkeley

The Ninth Annual Political Methodology Conference, organized by the methodology section of the American Political Science Association, will be held July 16–19, 1992 at Harvard University, Cambridge Massachusetts. Both faculty and graduate students are invited to apply to attend the conference.

Faculty who would like to participate should send a three to five page proposal describing the research they wish to present. Work in all areas of methodology will be considered such as time series and cross sectional methods, survey methodology, artificial intelligence, ecological inference, tests of formal models, content analysis, social network theory, multidimensional scaling and psychometric methods, comparative methodology, observational methods, statistical estimation and inference, experimental design, event history analysis and other areas. Papers may consider any substantive area in political science, but it is expected that there will be a strong connection to questions of interest to political scientists. Submissions are encouraged from younger scholars.

Advanced graduate students with interests in methodology will also be invited. These individuals will not be required to present a paper, although they are welcome to propose to do so. Graduate students should submit a letter explaining their interest in methodology and how attendance at the meetings will further their graduate work. A vita should accompany the application. In addition, graduate students should ask a faculty member to write a letter of recommendation including a brief description of the student's methodological training and research interests. Those students who are chosen to attend the meetings will be asked to prepare a month before the meetings a two to five page statement of their major research interests.

The National Science Foundation and Harvard University will cover most of the costs of attending the conference. A combined total of 35 faculty and graduate students will receive invitations. Other interested scholars are welcome to attend the sessions. However, only the 35 individuals who receive formal invitations will be reimbursed for travel expenses and provided with lodging and meals. Submissions should be sen to Henry E. Brady, Department of Political Science, University of California, Berkeley, California, 94720. The deadline for receipt of proposals in April 1, 1991. These submissions will be reviewed by a panel of the methodology section members. The selection of proposals will be announced on or about May 1, 1992.

A Call from the New Editor of *Political Analysis*

John R. Freeman
University of Minnesota

As the new editor of *Political Analysis*, I encourage all reader to submit their methodological article to me. I am interested in all types of methodology as well as in formal modeling. I am also interested in ideas people have for symposia and "Controversies" on methodological topics. Your work will be evaluated by referees who are more methodologically sophisticated than those who sometimes review your work for other journals. In addition, you will have the opportunity to publish more technical and/or ancillary material in *Political Analysis* than in other journals (e.g., proofs and results of sensitivity tests).

Information about manuscript format and style can be found in the opening pages of volumes 1 and 2 of *Political Analysis*. Submissions should be sent to: John R. Freeman, Editor, *Political Analysis*, Department of Political Science, 1414 Social Sciences Building, University of Minnesota, Minneapolis, MN 55455.

Contents of *Political Analysis*, Volume 3

The third volume of *Political Analysis* will be published shortly by the University of Michigan Press. *Political Analysis* is the journal of the Political Methodology Section of APSA and provides the primary outlet for research in political methodology. Special discount pricing for *PA* is available to members of the methodology section. You can also help support this annual volume by getting your university library to place a standing order for the journal. Here is an advance peek at the contents of the coming issue.

Thomas W. Gilligan: Performance of an Institutionalized Legislature.

John E. Jackson: Estimation of Models with Variable Coefficients.

Nathaniel Beck: Comparing Dynamic Specifications: The Case of Presidential Approval.

Elizabeth Theiss-Morse, Amy Fried, John L. Sullivan and Mary Dietz: Mixing Methods: A Multi-Stage Strategy for Studying Patriotism and Citizen Participation.

Jim Granato: An Agenda for Econometric Model Building.

Simon Jackman and Francis Vella: Electoral Redistricting and Endogenous Partisan Control.

New Officers for Political Methodology Section

According to the current bylaws, the Political Methodology Section has a five member Executive Committee composed of the President, Vice-President, Treasurer, the Editor of *Political Analysis* and an At-Large member. Nathaniel Beck of the University of California, San Diego, is the Executive Committee member with the greatest seniority, and he will complete his term as Treasurer in 1992. Henry Brady of the University of California, Berkeley and Larry Bartels of Princeton University were elected for two year terms as the President and Vice-President at the APSA meetings in September. Gary King of Harvard University (and a former editor of *The Political Methodologist*) was elected as the At-Large Member. John Freeman, a former President of the Methodology Section, took on the duties of editing *Political Analysis* in September. Political methodologists will note the political balance in this executive committee, with two specialists in time-series methods (Beck and Freeman), one specialist in psychometric methods (Brady), an expert on econometric techniques used in the analysis of cross-sectional data (Bartels), and a specialist in maximum likelihood and aggregate data analysis (King).

Political Methodology Explained

Philip Schrodt
University of Kansas

An anthropologist from a developing country came to the United States to study the culture of political science. After sitting in on faculty meetings and hiring decisions, attending the APSA convention, and talking with graduate students, departmental chairs and journal editors, the anthropologist reported the following:

> This culture seems most alien, but at least I understand the role of the political methodologists. They are the witch doctors. Everybody fears them; most people hate them. They are central to the rites of passage into the tribe. Nobody has the slightest idea what they do but witch doctors are thought to disrupt work from great distances—particularly cooking— so you keep them around for protection. And *everybody* agrees you only want one in the village.

Software News and Views

Aptech Systems, makers of Gauss, Markov and other software marvels, has moved to a new address: 23804 SE Kent-Kangley Rd., Maple Valley, WA 98038. Their phone number is 206-432-7855. Gauss is currently shipping in version 2.2. Markov, reviewed in the last issue of *TPM*, is also now shipping in version 1.0.

For those who missed the review, Markov is an easy to use statistical environment for Gauss. While not a complete statistical package, Markov includes modules for log-linear models, logit and probit, regression (including simultaneous equations) and basic statistics. As a convenient way to do most of what you need, it is very handy. Because Gauss (and Markov) make use of all the memory available on 386 and 486 machines (assuming you do the smart thing and buy the 386 version of Gauss), Markov is able to estimate large models that would stop most programs. I recently estimated some dichotomous probits with 15 independent variables and over 12,000 cases with no trouble at all. Because Gauss can be accessed directly from a Markov program, I was also easily able to write a two-stage probit program using Markov for most of the heavy lifting, while dipping into Gauss for some of the non-standard details. Markov is also available in a student version. Recommended! —*CHF*

The Political Methodologist
Department of Political Science
Washington University
One Brookings Drive
St. Louis, MO 63130

The Political Methodologist is the newsletter of the Political Methodology Section of the American Political Science Association. Copyright 1991, American Political Science Association. All rights reserved. We gratefully acknowledge the support of Washington University in helping to defray the editorial and production costs of the newsletter.

Subscriptions to *TPM* are free to members of the APSA's Methodology Section. Please contact APSA to join the section. Dues are $8.00 per year.

Submissions to *TPM* are welcome. Articles should be sent to the editor, Charles H. Franklin, Department of Political Science, Washington University, One Brookings Drive, St. Louis, MO 63130. If at all possible, please send articles via Bitnet or Internet to the editor at C38871CF@WUVMD. Submissions may also be made on diskette as plain ascii files. LaTeX format files are especially encouraged, though we can read most word processor files. The deadline for submissions for the next issue is March 15, 1992.