# Making the Most of Statistical Analyses: Improving Interpretation and Presentation

**Gary King** Harvard University
**Michael Tomz** Harvard University
**Jason Wittenberg** Harvard University

Social scientists rarely take full advantage of the information available in their statistical results. As a consequence, they miss opportunities to present quantities that are of greatest substantive interest for their research and express the appropriate degree of certainty about these quantities. In this article, we offer an approach, built on the technique of statistical simulation, to extract the currently overlooked information from any statistical method and to interpret and present it in a reader-friendly manner. Using this technique requires some expertise, which we try to provide herein, but its application should make the results of quantitative articles more informative and transparent. To illustrate our recommendations, we replicate the results of several published works, showing in each case how the authors' own conclusions can be expressed more sharply and informatively, and, without changing any data or statistical assumptions, how our approach reveals important new information about the research questions at hand. We also offer very easy-to-use software that implements our suggestions.

We show that social scientists often do not take full advantage of the information available in their statistical results and thus miss opportunities to present quantities that could shed the greatest light on their research questions. In this article we suggest an approach, built on the technique of statistical simulation, to extract the currently overlooked information and present it in a reader-friendly manner. More specifically, we show how to convert the raw results of any statistical procedure into expressions that (1) convey numerically precise estimates of the quantities of greatest substantive interest, (2) include reasonable measures of uncertainty about those estimates, and (3) require little specialized knowledge to understand.

The following simple statement satisfies our criteria: "Other things being equal, an additional year of education would increase your annual income by $1,500 on average, plus or minus about $500." Any smart high school student would understand that sentence, no matter how sophisticated the statistical model and powerful the computers used to produce it. The sentence is substantively informative because it conveys a key quantity of interest in terms the reader wants to know. At the same time, the sentence indicates how uncertain the researcher is about the estimated quantity of interest. Inferences are never certain, so any honest presentation of statistical results must include some qualifier, such as "plus or minus $500" in the present example.

Gary King is Professor of Government, Harvard University, Littauer Center North Yard, Cambridge, MA 02138 (king@harvard.edu). Michael Tomz is a Ph.D. candidate, Department of Government, Harvard University, Littauer Center North Yard, Cambridge, MA 02138 (tomz@fas.harvard. edu). Jason Wittenberg is a fellow of the Weatherhead Center for International Affairs, Cambridge Street, Cambridge, MA 02138 (jwittenb @latte.harvard.edu).

Our computer program, "CLARIFY: Software for Interpreting and Presenting Statistical Results," designed to implement the methods described in this article, is available at http://GKing.Harvard.Edu, and won the Okidata Best Research Software Award for 1999. We thank Bruce Bueno de Mesquita, Jorge Domínguez, Geoff Garrett, Jay McCann, and Randy Siverson for their data; Jim Alt, Alison Alter, Steve Ansolabehere, Marc Busch, Nick Cox, Jorge Domínguez, Jeff Gill, Michael Meffert, Jonathan Nagler, Ken Scheve, Brian Silver, Richard Tucker, and Steve Voss for their very helpful comments; and U.S./National Science Foundation (SBR-9729884), the Centers for Disease Control and Prevention (Division of Diabetes Translation), the National Institutes of Aging, the World Health Organization, and the Global Forum for Health Research for research support.

*American Journal of Political Science*, Vol. 44, No. 2, April 2000, Pp. 341–355

In contrast, bad intepretations are substantively ambiguous and filled with methodological jargon: "the coefficient on education was statistically significant at the 0.05 level." Descriptions like this are very common in social science, but students, public officials, and scholars should not need to understand phrases like "coefficient," "statistically significant," and "the 0.05 level" to learn from the research. Moreover, even statistically savvy readers should complain that the sentence does not convey the key quantity of interest: how much higher the starting salary would be if the student attended college for an extra year.

Our suggested approach can help researchers do better in three ways. First, and most importantly, it can extract new quantities of interest from standard statistical models, thereby enriching the substance of social science research. Second, our approach allows scholars to assess the uncertainty surrounding any quantity of interest, so it should improve the candor and realism of statistical discourse about politics. Finally, our method can convert raw statistical results into results that everyone, regardless of statistical training, can comprehend. The examples in this article should make all three benefits apparent.

Most of this article describes our simulation-based approach to interpreting and presenting statistical results. In many situations, what we do via simulation can also be done by direct mathematical analysis or other computationally-intensive techniques, and we discuss these approaches as well. To assist researchers in implementing our suggestions, we have developed an easy-to-use, public domain software package called CLARIFY, which we describe in the appendix.

## The Problem of Statistical Interpretation

We aim to interpret the raw results from any member of a very general class of statistical models, which we summarize with two equations:

$$Y_i \sim f(\theta_i, \alpha), \quad \theta_i = g(X_i, \beta). \tag{1}$$

The first equation describes the *stochastic component* of the statistical model: the probability density (or mass) function that generates the dependent variable $Y_i$ ($i = 1, \ldots, n$) as a random draw from the probability density $f(\theta_i, \alpha)$. Some characteristics of this function vary from one observation to the next, while others remain constant across all the $i$'s. We represent the varying characteristics with the parameter vector $\theta_i$ and relegate nonvarying features to the ancillary parameter matrix $\alpha$. The second equation gives the systematic component of the model; it indicates how $\theta_i$ changes across observations, depending on values of explanatory variables (typically including a constant) in the $1 \times k$ vector $X_i$ and effect parameters in the $k \times 1$ vector $\beta$. The functional form $g(\cdot,\cdot)$, sometimes called the link function, specifies how the explanatory variables and effect parameters get translated into $\theta_i$.

One member of this general class is a linear-normal regression model, otherwise known as least-squares regression. To see this, let $f(\cdot,\cdot)$ be the normal distribution $N(\cdot,\cdot)$; set the main parameter vector to the scalar mean $\theta_i = E(Y_i) = \mu_i$; and assume that the ancilliary parameter matrix is the scalar homoskedastic variance $\alpha = V(Y_i) = \sigma^2$. Finally, set the systematic component to the linear form $g(X_i,\beta) = X_i\beta = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + \cdots$. The result is familiar:

$$Y_i \sim N(\mu_i,\sigma^2), \qquad \mu_i = X_i\beta \tag{2}$$

Similarly, one could write a logit model by expressing the stochastic component as a Bernoulli distribution with main parameter $\pi_i = \Pr(Y_i = 1)$—no ancillary parameter is necessary—and setting the systematic component to the logistic form:

$$Y_i \sim \text{Bernoulli}(\pi_i), \qquad \pi_i = \frac{1}{1 + e^{-X_i\beta}} \tag{3}$$

Equation 1 also includes as special cases nearly every other statistical model in the social sciences, including multiple-equation models in which $Y_i$ is a vector, as well as specifications for which the probability distribution, functional form, or matrix of explanatory variables is estimated rather than assumed to be known.

Having estimated the statistical model, many researchers stop after a cursory look at the signs and "statistical significance" of the effect parameters. This approach obviously fails to meet our criteria for meaningful statistical communication since, for many nonlinear models, $\hat{\beta}$ and $\hat{\alpha}$ are difficult to interpret and only indirectly related to the substantive issues that motivated the research (Cain and Watts 1970; Blalock 1967). Instead of publishing the effect coefficients and ancillary parameters, researchers should calculate and present quantities of direct substantive interest.

Some researchers go a step farther by computing derivatives, fitted values, and first differences (Long 1997; King 1989, subsection 5.2) which do convey numerically precise estimates of interesting quantities and require little specialized knowledge to understand. Even these approaches are inadequate, however, because they ignore two forms of uncertainty. *Estimation uncertainty* arises

from not knowing $\beta$ and $\alpha$ perfectly, an unavoidable consequence of having fewer than an infinite number of observations. Researchers often acknowledge this uncertainty by reporting standard errors or *t*-statistics, but they overlook it when computing quantities of interest. Since $\hat{\beta}$ and $\hat{\alpha}$ are uncertain, any calculations—including derivatives, fitted values, and first differences—based on those parameter estimates must also be uncertain, a fact that almost no scholars take into account. A second form of variability, the *fundamental uncertainty* represented by the stochastic component (the distribution $f$) in Equation 1, results from innumerable chance events such as weather or illness that may influence $Y$ but are not included in $X$. Even if we knew the exact values of the parameters (thereby eliminating estimation uncertainty), fundamental uncertainty would prevent us from predicting $Y$ without error. Our methods for computing quantities of interest must account for both types of uncertainty.

## Simulation-Based Approaches to Interpretation

We recommend statistical simulation as an easy method of computing quantities of interest and their uncertainties. Simulation can also help researchers understand the entire statistical model, take full advantage of the parameter estimates, and convey findings in a reader-friendly manner (see Fair 1980; Tanner 1996; Stern 1997).

### What Is Statistical Simulation?

Statistical simulation uses the logic of survey sampling to approximate complicated mathematical calculations. In survey research, we learn about a population by taking a random sample from it. We use the sample to estimate a feature of the population, such as its mean or its variance, and our estimates become more precise as we increase the sample size, $n$. Simulation follows a similar logic but teaches us about probability distributions, rather than populations. We learn about a distribution by simulating (drawing random numbers) from it and using the draws to approximate some feature of the distribution. The approximation becomes more accurate as we increase the number of draws, $M$. Thus, simulation enables us to approximate any feature of a probability distribution without resorting to advanced mathematics.

For instance, we could compute the mean of a probability distribution $P(y)$ by taking the integral $E(Y) = \int_{-\infty}^{\infty} yP(y)dy$, which is not always the most pleasant of experiences! Alternatively, we could approximate the mean through simulation by drawing many random numbers from $P(y)$ and computing their average. If we were interested in the theoretical variance of $Y$, we could calculate the sample variance of a large number of random draws, and if we wanted the probability that $Y > 0.8$, we could count the fraction of draws that exceeded 0.8. Likewise, we could find a 95-percent confidence interval for a *function* of $Y$ by drawing 1000 values of $Y$, computing the function for each draw, sorting the transformed draws from lowest to highest and taking the 25th and 976th values. We could even approximate the entire distribution of, say, $\sqrt{Y}$, by plotting a histogram of the square roots of a large number of simulations of $Y$.

Approximations can be computed to *any* desired degree of precision by increasing the number of simulations ($M$), which is analagous to boosting the number of observations in survey sampling. Assessing the precision of the approximation is simple: run the same procedure, with the same number of simulations, repeatedly. If the answer remains the same to within four decimal points across the repetitions, that is how accurate the approximation is. If more accuracy is needed, raise the number of simulations and try again. Nothing is lost by simulation—except a bit of computer time—and much is gained in ease of use.

### Simulating the Parameters

We now explain how researchers can use simulation to compute quantities of interest and account for uncertainty. The first step involves simulating the main and ancillary parameters. Recall that the parameter estimates $\hat{\beta}$ and $\hat{\alpha}$ are never certain because our samples are finite. To capture this estimation uncertainty, we draw many plausible sets of parameters from their posterior or sampling distribution. Some draws will be smaller or larger than $\hat{\beta}$ and $\hat{\alpha}$, reflecting our uncertainty about the exact value of the parameters, but all will be consistent with the data and statistical model.

To simulate the parameters, we need the point estimates and the variance-covariance matrix of the estimates, which most statistical packages will report on request. We denote $\hat{\gamma}$ as the vector produced by stacking $\hat{\beta}$ on top of $\hat{\alpha}$. More formally, $\hat{\gamma} = \text{vec}(\hat{\beta}, \hat{\alpha})$, where "vec" stacks the unique elements of $\hat{\beta}$ and $\hat{\alpha}$ in a column vector. Let $\hat{\text{V}}(\hat{\gamma})$ designate the variance matrix associated with these estimates. The central limit theorem tells us that with a large enough sample and bounded variance, we can randomly draw (simulate) the parameters from a

multivariate normal distribution with mean equal to $\hat{\gamma}$ and variance equal to $\hat{V}(\hat{\gamma})$.[1] Using our notation,

$$\tilde{\gamma} \sim N\left(\hat{\gamma}, \hat{V}(\hat{\gamma})\right). \quad (4)$$

Thus, we can obtain *one* simulation of $\gamma$ by following these steps:

1. Estimate the model by running the usual software program (which usually maximizes a likelihood function), and record the point estimates $\hat{\gamma}$ and variance matrix $\hat{V}(\hat{\gamma})$.
2. Draw one value of the vector $\gamma$ from the multivariate normal distribution in Equation 4. Denote the $\tilde{\gamma} = \text{vec}(\tilde{\beta}, \tilde{\alpha})$.

Repeat the second step, say, $M = 1000$ times to obtain 1000 draws of the main and ancillary parameters.

If we knew the elements of $\gamma$ perfectly, the sets of draws would all be identical; the less information we have about $\gamma$ (due to larger elements in the variance matrix), the more the draws will differ from each other. The specific pattern of variation summarizes all knowledge about the parameters that we can obtain from the statistical procedure. We still need to translate $\gamma$ into substantively interesting quantities, but now that we have summarized all knowledge about $\gamma$ we are well positioned to make the translation. In the next three subsections, we describe algorithms for converting the simulated parameters into predicted values, expected values, and first differences.

## Predicted Values

Our task is to draw one value of $Y$ conditional on one chosen value of each explanatory variable, which we represent with the vector $X_c$. Denote the simulated $\theta$ as $\tilde{\theta}_c$ and the corresponding Y as $\tilde{Y}_c$, a simulated predicted value. Predicted values come in many varieties, depending on the kind of $X$-values used. For instance, $X_c$ may correspond to the future (in which case $\tilde{Y}_c$ is a simulated forecast), a real situation described by observed data (such that $\tilde{Y}_c$ is a simulated predicted value), or a hypothetical situation not necessarily in the future (making $\tilde{Y}_c$ a simulated counterfactual predicted value). None of these is equivalent to the expected value ($\hat{Y}$) in a linear regression, which we discuss in the following subsection.

To simulate *one* predicted value, follow these steps:

1. Using the algorithm in the previous subsection, draw one value of the vector $\tilde{\gamma} = \text{vec}(\tilde{\beta}, \tilde{\alpha})$.
2. Decide which kind of predicted value you wish to compute, and on that basis choose one value for each explanatory variable. Denote the vector of such values $X_c$.
3. Taking the simulated effect coefficients from the top portion of $\tilde{\gamma}$, compute $\tilde{\theta}_c = g(X_c, \tilde{\beta})$, where $g(\cdot,\cdot)$ is the systematic component of the statistical model.
4. Simulate the outcome variable $\tilde{Y}_c$ by taking a random draw from $f(\tilde{\theta}_c, \tilde{\alpha})$, the stochastic component of the statistical model.

Repeat this algorithm, say, $M = 1000$ times, to produce 1000 predicted values, thereby approximating the entire probability distribution of $Y_c$. From these simulations the researcher can compute not only the average predicted value but also measures of uncertainty around the average. The predicted value will be expressed in the same metric as the dependent variable, so it should require little specialized knowledge to understand.

## Expected Values

Depending on the issue being studied, the *expected* or mean value of the dependent variable may be more interesting than a predicted value. The difference is subtle but important. A predicted value contains both fundamental and estimation uncertainty, whereas an expected value averages over the fundamental variability arising from sheer randomness in the world, leaving only the estimation uncertainty caused by not having an infinite number of observations. Thus, predicted values have a larger variance than expected values, even though the average should be nearly the same in both cases.[2]

When choosing between these two quantities of interest, researchers should reflect on the importance of fundamental uncertainty for the conclusions they are drawing. In certain applications, such as forecasting the actual result of an election or predicting next month's foreign exchange rate, scholars and politicians—as well as investors—want to know not only the expected outcome, but also how far the outcome could deviate from expectation due to unmodeled random factors. Here, a

[1]This distributional statement is a shorthand summary of the Bayesian, likelihood, and Neyman-Pearson theories of statistical inference. The interpretive differences among these theories (such as whether $\theta$ or $\hat{\theta}$ is the random variable) are important but need not concern us here, as our approach can usually be employed with any of these and most other theories of inference (see Barnett 1982).

[2]In linear models, the average predicted value is identical to the expected value. For nonlinear cases, the two can differ but are often close if the nonlinearity is not severe.

predicted value seems most appropriate. For other applications, the researcher may want to highlight the average effect of a particular explanatory variable, so an expected value would be the best choice.

We now offer an algorithm for creating *one* simulation of an expected value:

1. Following the procedure for simulating the parameters, draw one value of the vector $\tilde{\gamma} = \text{vec}(\tilde{\beta}, \tilde{\alpha})$.
2. Choose one value for each explanatory variable and denote the vector of values as $X_c$.
3. Taking the simulated effect coefficients from the top portion of $\tilde{\gamma}$, compute $\tilde{\theta}_c = g(X_c, \tilde{\beta})$, where $g(\cdot,\cdot)$ is the systematic component of the statistical model.
4. Draw $m$ values of the outcome variable $\tilde{Y}_c^{(k)}$ ($k = 1, \ldots, m$) from the stochastic component $f(\tilde{\theta}_c, \tilde{\alpha})$. This step simulates fundamental uncertainty.
5. Average over the fundamental uncertainty by calculating the the mean of the $m$ simulations to yield one simulated expected value $\tilde{E}(Y_c) = \sum_{k=1}^{m} \tilde{Y}_c^{(k)} \big/ m$.

When $m = 1$, this algorithm reduces to the one for predicted values. If $m$ is a larger number, Step 4 accurately portrays the fundamental variability, which Step 5 averages away to produce an expected value. The larger the value of $m$, the more successful the algorithm will be in purging $\tilde{E}(Y_c)$ of any fundamental uncertainty.

To generate 1000 simulations of the expected value, repeat the entire algorithm $M = 1000$ times for some fixed value of $m$. The resulting expected values will differ from each other due to estimation uncertainty, since each expected value will correspond to a different $\tilde{\gamma}$. These $M$ simulations will approximate the full probability distribution of $E(Y_c)$, enabling the researcher to compute averages, standard errors, confidence intervals, and almost anything else desired.

The algorithm works in all cases but involves some approximation error, which we can reduce by setting both $m$ and $M$ sufficiently high. For some statistical models, there is a shortcut that curtails both computation time and approximation error. Whenever $E(Y_c) = \theta_c$, the researcher can skip steps 4–5 of the expected value algorithm, since steps 1–3 suffice to simulate one expected value. This shortcut is appropriate for the linear-normal and logit models in Equations 2 and 3.

## First Differences

A first difference is the difference between two expected, rather than predicted, values. To simulate a first difference, researchers need only run steps 2–5 of the expected value algorithm twice, using different settings for the explanatory variables.

For instance, to simulate a first difference for the first explanatory variable, set the values for all explanatory variables except the first at their means and fix the first one at its starting point. Denote this vector of starting values for the explanatory variables as $X_s$ and run the expected value algorithm once to generate $\tilde{E}(Y_s)$, the average value of $Y$ conditional on $X_s$. Next change the value of the first explanatory variable to its ending point, leaving the others at their means as before. Denote the new vector as $X_e$ and rerun the algorithm to get $\tilde{E}(Y_e)$, the mean of $Y$ conditional on $X_e$. The first difference is simply $\tilde{E}(Y_e) - \tilde{E}(Y_s)$. Repeat the first difference algorithm, say, $M = 1000$ times to approximate the distribution of first differences. Average the simulated values to obtain a point estimate, compute the standard deviation to obtain a standard error, or sort the values to approximate a confidence interval.

We previously discussed expected values of $Y$, and until now this section has considered first differences based on only this type of expected value. Different expectations, such as $\Pr(Y = 3)$ in an ordered-probit model, may also be of interest. For these cases, the expected value algorithm would need to be modified slightly. We have made the necessary modifications in CLARIFY, the software package described in the appendix, which allows researchers to calculate a wide variety of expected values and first differences, as well as predicted values and other quantities of interest.

The algorithms in this article do not require new assumptions; rather, they rest on foundations that have become standard in the social sciences. In particular, we assume that the statistical model is identified and correctly specified (with the appropriate explanatory variables and functional form), which allows us to focus on interpreting and presenting the final results. We also assume that the central limit theorem holds sufficiently for the available sample size, such that the sampling distribution of parameters (not the stochastic component) can be described by a normal distribution.[3] Although we focus on asymptotic results, as do the vast majority of the applied researchers using nonlinear models, simulation works with finite sample distributions, which are preferable when feasible. In short, our algorithms work whenever the usual assumptions work.

## Alternative Approaches

In this section, we discuss several other techniques for generating quantities of interest and measuring the uncertainty around them. These approaches can be valuable

[3]From a Bayesian perspective, we exclude unusual cases where a flat prior generates an improper posterior.

complements to simulation, because they provide important mathematical intuition or, in some cases, enable finite sample approximations. We briefly summarize some of the leading computer-intensive and analytical alternatives to simulation.

**Computer-intensive alternatives.** Our version of simulation is not the only computer-intensive technique for obtaining quantities of interest and measures of uncertainty. Fully Bayesian methods, using Markov-Chain Monte Carlo techniques, are more powerful than our algorithms because they allow researchers to draw from the exact finite-sample distribution, instead of relying on the central limit theorem to justify an asymptotic normal approximation (Carlin and Louis 1996). Unfortunately these methods remain difficult to use, particularly since statisticians still disagree about appropriate criteria for determining when a Markov chain has converged in distribution to the true posterior (Cowles and Carlin 1996; Kass et al. 1998). Nonetheless, this field has shown remarkable progress over the last decade and is well worth monitoring by political scientists.

Another useful alternative is bootstrapping, a nonparametric approach that relies on the logic of *re*-sampling to approximate the distribution of parameters (Mooney and Duval 1993; Mooney 1996). In theory, the sampling distribution of $\hat{\gamma}$ can be viewed as a histogram of an infinite number of $\hat{\gamma}$'s, each estimated from a different sample of size $n$ from the same population. Bootstrapping mimicks this process by drawing many subsamples (with replacement) from the original sample, estimating $\hat{\gamma}$ for each subsample, and then constructing a histogram of the various $\hat{\gamma}$'s. Bootstrapping has many advantages. It does not require strong distributional assumptions, and Monte Carlo studies have demonstrated that it has superior small sample properties for some problems. It also has the advantage of not requiring strong parametric distributional assumptions. Programming a bootstrapped estimator is not difficult, although commercial software packages have not been quick to adopt this method. The weakness of bootstrapping is that it gives biased estimates for certain quantities of interest, such as $\max(Y)$.

For both Bayesian methods and bootstrapping, all of the methods of interpretation we discuss in this article can be used directly. The only change is that instead of drawing the parameters from the multivariate normal in Equation 4, we would use MCMC-based simulation or bootstrapping. Even our software, CLARIFY, could be used without additional programming.

Like our method, MCMC and bootstrapping generate simulations of the parameters. In cases where the parameters are not of intrinsic interest, researchers must convert them into quantities such as predicted values, expected values, and first differences. The algorithms above show how to make the conversion and are therefore essential supplements. Indeed, our software, CLARIFY, could easily be modified to interpret the parameters generated by these alternative approaches.

**Analytical approaches.** The main analytical (mathematical) alternative to simulation is the delta method, which uses the tools of calculus to approximate nonlinear functions of random variables (van der Vaart 1998). Suppose that we are interested in the mean and variance of $\theta = g(X_c, \beta)$, where $g$ is a nonlinear function. Assuming that $g$ is approximately linear in a range where $\beta$ has high probability, then a Taylor-series expansion of $g$ about $\hat{\beta}$ is often reasonable. To the first order, $\theta \approx g(\hat{\beta}) + g'(\hat{\beta})(\beta - \hat{\beta})$, where $g'(a) = \partial g(a)/\partial a$. As a result, the maximum-likelihood estimate of $\theta$ is approximately $g(\hat{\beta})$, and its variance is approximately $g'(\hat{\beta})V(\hat{\beta})g'(\hat{\beta})'$. For example, in the exponential Poisson regression model (King 1989, Chapter 5), where $Y$ is Poisson with mean $\lambda = E(Y|X) = e^{X_i\beta}$, suppose we wish to compute the expected number of events given $X = X_0$. In this case, the maximum likelihood estimate of the expected number of events is $g(\hat{\beta}) = e^{X_0\hat{\beta}}$ and its variance is $(X_0 e^{X_0\hat{\beta}})\hat{V}(\hat{\beta})(X_0 e^{X_0\hat{\beta}})'$. Note that this maximum-likelihood estimate still does not reflect the uncertainty in $\hat{\beta}$, as done automatically by simulation and the other computationally intensive methods. To incorporate this additional uncertainty requires another level of mathematical complexity in that we must now approximate the integral $\int e^{X_0\hat{\beta}} P(\hat{\beta}) d\hat{\beta}$ and its variance. A detailed example is given by King and Zeng (1999) in the case of logistic regression.

Despite its utility for increasing computing speed and revealing statistical intuition through mathematical analysis, the delta method suffers from two shortcomings that simulation can help overcome. First, the method is technically demanding, since it requires researchers to calculate derivatives and compute the moments of linearized functions.[4] Thus, it is not surprising that most scholars do not use the delta method, even when they appreciate the importance of reporting uncertainty. Second, the Taylor series used in the delta method only approximates a nonlinear form. Although researchers can sometimes improve the approximation with additional terms in the Taylor series, this can be difficult, and find-

[4] When $g$ is linear there is obviously no need for a linearizing approximation; an exact analytical solution exists for the mean and variance of many quantities of interest that we described earlier. Simulation produces the same answer, however, and requires less mathematical proficiency.

ing estimates of the additional terms is often impossible. In practice most researchers stop after expanding the series to the first or second order, which can compromise the accuracy of the approximation. With simulation one can achieve an arbitrarily high degree of precision simply by increasing $M$ and letting the computer run longer.

Several general arguments weigh in favor of simulation. First, there is a simulation-based alternative to nearly every analytical method of computing quantities of interest and conducting statistical tests, but the reverse is not true (Noreen 1989). Thus, simulation can provide accurate answers even when no analytical solutions exist. Second, simulation enjoys an important pedagogical advantage. Studies have shown that, no matter how well analytical methods are taught, students get the right answer far more often via simulation (Simon, Atkinson, and Shevokas 1976). One scholar has even offered a $5,000 reward for anyone who can demonstrate the superiority of teaching analytical methods, but so far no one has managed to earn the prize (Simon 1992). Of course, there are advantages to the insight the mathematics underlying the delta method can reveal and so, when feasible, we encourage researchers to learn both simulation and analytical methods.

## Tricks of the Trade

The algorithms in the previous section apply to all statistical models, but they can be made to work better by following a few tricks and avoiding some common misunderstandings.

### Tricks for Simulating Parameters

Statistical programs usually report standard errors for parameter estimates, but accurate simulation requires the full variance matrix $\hat{V}(\hat{\gamma})$. The diagonal elements of $\hat{V}(\hat{\gamma})$ contain the squared standard errors, while the off-diagonal elements express the covariances between one parameter estimate and another in repeated draws from the same probability distribution. Simulating each parameter independently would be incorrect, because this procedure would miss the covariances among parameters. Nearly all good statistical packages can report the full variance matrix, but most require researchers to request it explicitly by setting an option or a global. The software described in the appendix fetches the variance matrix automatically.

One common mistake is to exclude some parameters when drawing from the multivariate normal distribution. Parameters have different logical statuses, such as the effect parameters $\beta$ versus the ancillary parameters $\alpha$, but our algorithms do not need to distinguish between the two: both are uncertain and should be simulated, even if only one proves useful in later calculations. It may be possible to accelerate our algorithms by excluding certain parameters from the simulation stage, but for the vast majority of applications these tricks are unnecessary and could lead to errors. Researchers usually will risk fewer mistakes by following, without deviation, our algorithm for simulating the parameters.

In some statistical models, the elements of $\gamma$ are orthogonal, so software packages provide separate variance matrices for each set. When implementing the algorithm for simulating the parameters, researchers may want to create a bloc diagonal matrix by placing the separately estimated variance matrices on the diagonal and inserting zeros everywhere else. Obviously, if the subsets of $\gamma$ truly are orthogonal, equivalent draws from the two sets can be made from independent multivariate normal distributions, but it may be easier to work with a single sampling distribution.

Researchers should reparameterize the elements of $\gamma$ to increase the likelihood that the asymptotic multivariate normal approximation will hold in finite samples. In general, all parameters should be reparameterized unless they are already unbounded and logically symmetric, as a Normal must be. For instance, a variance parameter like $\sigma^2$ must be greater than zero, so it will pay to reparameterize by using an expression like $\sigma^2 = e^{\eta}$. This allows researchers to estimate $\eta$, which is on the scale from $-\infty$ to $\infty$, as one element of $\gamma$, which is assumed to be multivariate normal. When making reparameterizations, of course, we add an extra step to the algorithm for simulating the parameters: after drawing $\gamma$ from the multivariate normal, we reparameterize back to the original scale by computing $\tilde{\sigma}^2 = e^{\tilde{\eta}}$.[5]

Several other reparameterizations may come in handy. A correlation parameter $\rho$, ranging from –1 to 1, can be reparameterized to $\eta$ (reusing the same symbol) on an unbounded scale with the inverse of Fisher's Z transformation: $\rho = (e^{2\eta} - 1)/(e^{2\eta} + 1)$. Likewise, a parameter representing a probability $\pi$ can be made

[5]Reparameterization also makes likelihood-maximization algorithms easier to use by avoiding problems caused by the optimization procedure choosing inadmissable parameter values (which often result in the program terminating abnormally because of attempts to divide by zero or logging negative numbers). Since maximum likelihood estimates are invariant to reparameterization, the reparameterization has no effect except on the finite sample distribution around the point estimate. For example, estimating $\hat{\sigma}^2$ directly gives the same estimate as estimating $\hat{\eta}$ and transforming to $\hat{\sigma}^2$ by using $\hat{\sigma}^2 = e^{\hat{\eta}}$.

unbounded using the logistic transformation, $\pi = [1 + e^{-\eta}]^{-1}$. These and other tricks should enhance the effectiveness of simulating the parameters.

### Tricks for Simulating Quantities of Interest

When converting the simulated parameters into quantities of interest, it is safest to simulate $Y$ and use this as a basis for obtaining other quantities. This rule is equivalent to incorporating all simulated parameters—and thus all information from the statistical model—into the calculations. Of course, some shortcuts do exist. We have already mentioned that, in a logit model, one can obtain $\tilde{E}(Y)$ by stopping with $\tilde{\pi}$, since drawing dichotomous $Y$'s and averaging would yield exactly $\tilde{\pi}$. If one is not sure, though, it is helpful to continue until one has simulations of the outcome variable.

If some function of $Y$, such as $\ln(Y)$, is used as the dependent variable during the estimation stage, the researcher can simulate $\ln(Y)$ and then apply the inverse function $\exp(\ln(Y))$ to reveal $Y$. We adopt this procedure in the first example, below, where we estimate a log-log regression model. This sequence of simulation and transformation is crucial, since the usual procedure of calculating $E(\widehat{\ln(Y)}) = \hat{\mu}$ without simulation and then exponentiating gives the wrong answer: $\exp(\hat{\mu}) \neq \hat{Y}$. With simulation, both $Y$ and $E(Y)$ can be computed easily, regardless of the scale that the researcher used during the estimation stage.

Researchers should assess the precision of any simulated quantity by repeating the entire algorithm and seeing if anything of substantive importance changes. If something does change, increase the number of simulations ($M$ and, in the case of expected values, $m$) and try again. In certain instances—particularly when the researcher has misspecified a nonlinear statistical model—the number of simulations required to approximate an expected value accurately may be larger than normal. Numerical estimates should be reported to the correct level of precision, so for instance if repeated runs with the same number of simulations produce an estimate that changes only in the fourth decimal point, then—assuming this is sufficient for substantive purposes—the number reported should be rounded to two or three decimal points.

The simulation procedures given in this article can be used to compute virtually all quantities that might be of interest for nearly all parametric statistical models that scholars might wish to interpret. As such, they can be considered canonical methods of simulation. Numerous other simulation algorithms are available, however, in the context of specific models. When these alternatives could speed the approximation or make it more accurate for a fixed number of simulations, they should be incorporated into computer programs for general use. In some cases, analytical computations are also possible and can get speedier results. But our algorithms provide social scientists with all they need to understand the fundamental concepts: which quantities are being computed and how the computation is, or could be, done. Moreover, as long as $M$ and $m$ are large enough, these and all other correct algorithms will give identical answers.

## Empirical Examples

To illustrate how our algorithms work in practice, we include replications or extensions of five empirical works. Instead of choosing the most egregious, we choose a large number of the best works, from our most prestigious journals and presses, written by some of our most distinguished authors. Within this group, we eliminated the many publications we were unable to replicate and then picked five to illustrate a diverse range of models and interpretative issues. The procedures for model interpretation in all five were exemplary. If we all followed their examples, reporting practices in the discipline would be greatly improved. For each article, we describe the substantive problem posed and statistical model chosen; we also accept rather than evaluate their statistical procedures, even though in some cases the methods could be improved. We then detail how the authors interpreted their results and demonstrate how our procedures advance this state of the art.

### Linear Regression

Following Tufte (1974), we estimated a log-log regression model of the size of government in the U.S. states. Our dependent variable, $Y_i$, was the natural log of the number of people (measured in 1000s) that the state government employed on a full-time basis in 1990. Tufte was interested (for a pedagogical example) in whether $Y_i$ would increase with state population; but consider another hypothesis that may be of more interest to political scientists: the number of employees might depend on the proportion of Democrats in the state legislature, since Democrats are reputed to favor bigger government than Republicans, even after adjusting for state population. Thus, our two main explanatory variables were the log of state population $P_i$ in 1000s and the logged proportion of lower-house legislators who identified themselves as Democrats $D_i$.

We applied the predicted value algorithm to predict the number of government employees in a state with six million people and an 80 percent Democratic house. First, we used the statistical software described in the appendix to estimate the log-linear model and simulate one set of values for the effect coefficients $(\tilde{\beta})$ and the ancillary parameter $(\tilde{\sigma})$. Next, we set the main explanatory variables at $P_c = \ln(6000)$ and $D_c = \ln(0.8)$, so we could construct $X_c$ and compute $\tilde{\theta}_c = X_c\tilde{\beta}$. We then drew one value of $\tilde{Y}_c$ from the normal distribution $N(\tilde{\theta}_c, \sigma^2)$. Finally, we calculated $\exp(\tilde{Y}_c)$ to transform our simulated value into the actual number of government employees, a quantity that seemed more understandable than its natural logarithm. By repeating this process $M = 1000$ times, we generated 1000 predicted values, which we sorted from lowest to highest. The numbers in the 25th and the 976th positions represented the upper and lower bounds of a 95-percent confidence interval. Thus, we predicted with 95-percent confidence that the state government would employ between 73,000 and 149,000 people. Our best guess was 106,000 full-time employees, the average of the predicted values.

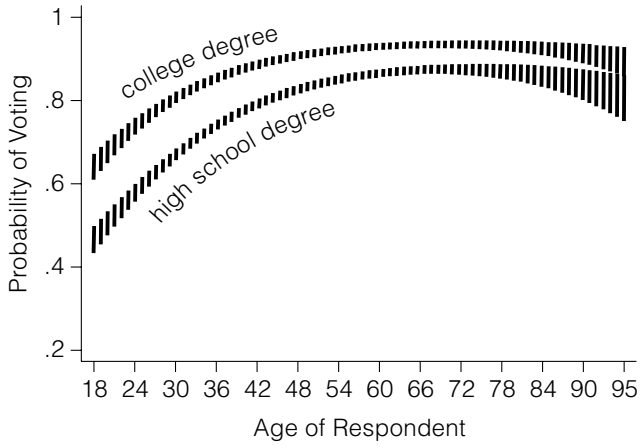
We also calculated some expected values and first differences and found that increasing Democratic control from half to two-thirds of the lower house tended to raise state government employment by 7,000 people on average. The 95-percent confidence interval around this first difference ranged from 3,000 to 12,000 full-time employees. Our result may be worth following up, since, to the best of our knowledge, researchers have not addressed this relationship in the state-politics literature.

## Logit Models

The algorithms in the third section can also help researchers interpret the results of a logit model. Our example draws on the work of Rosenstone and Hansen (1993), who sought to explain why some individuals are more likely than others to vote in U.S. presidential elections. Following Rosenstone and Hanson, we pooled data from every National Election Study that was conducted during a presidential election year. Our dependent variable, $Y_i$, was coded 1 if the respondent reported voting in the presidential election and 0 otherwise.

For expository purposes we focus on a few demographic variables that Rosenstone and Hanson emphasized: Age $(A_i)$ and Education $(E_i)$ in years, Income $(I_i)$ in 10,000s of dollars, and Race (coded $R_i = 1$ for whites and 0 otherwise). We also include a quadratic term to test the hypothesis that turnout rises with age until the respondent nears retirement, when the tendency reverses itself. Thus, our set of explanatory variables is $X_i = \{\mathbf{1}, A_i, A_i^2, E_i, I_i, R_i\}$, where $\mathbf{1}$ is a constant and $A_i^2$ is the quadratic term.

**Figure 1 Probability of Voting by Age**




Vertical bars indicate 99-percent confidence intervals

In our logit model, the probability of voting in a presidential election is $E(Y_i) = \pi_i$, an intuitive quantity of interest. We estimated this probability, and the uncertainty surrounding it, for two different levels of education and across the entire range of age, while holding other variables at their means. In each case, we repeated the expected value algorithm $M = 1000$ times to approximate a 99-percent confidence interval around the probability of voting. The results appear in Figure 1, which illustrates the conclusions of Rosenstone and Hansen quite sharply: the probability of voting rises steadily to a plateau between the ages of 45 and 65, and then tapers downward through the retirement years. The figure also reveals that uncertainty associated with the expected value is greatest at the two extremes of age: the vertical bars, which represent 99-percent confidence intervals, are longest when the respondent is very young or old.[6]

## A Time-Series Cross-Sectional Model

We also used our algorithms to interpret the results of a time-series cross-sectional model. Conventional wisdom holds that the globalization of markets has compelled governments to slash public spending, but a new book by Garrett (1998) offers evidence to the contrary. Where strong leftist parties and encompassing trade unions coincide, Garrett argues, globalization leads to *greater*

[6] The confidence intervals are quite narrow, because the large number of observations ($N = 15{,}837$) eliminated most of the estimation uncertainty.

government spending as a percentage of GDP, whereas the opposite occurs in countries where the left and labor are weak.

To support his argument, Garrett constructed a panel of economic and political variables, measured annually, for fourteen industrial democracies during the period 1966–1990. He then estimated a linear-normal regression model where the dependent variable, $Y_i$, is government spending as a percentage of GDP for each country-year in the data set. The three key explanatory variables were Capital mobility, $C_i$ (higher values indicate fewer government restrictions on cross-border financial flows), Trade, $T_i$ (larger values mean more foreign trade as a percentage of GDP), and Left-labor power, $L_i$ (higher scores denote a stronger combination of leftist parties and labor unions).[7]

To interpret his results, Garrett computed

> a series of counterfactual estimates of government spending under different constellations of domestic political conditions and integration into global markets. This was done by set ting all the other variables in the regression equations equal to their mean levels and multiplying these means by their corresponding coefficients, and then by examining the counterfactual impact of various combinations of left-labor power and globalization. . . . (1998, 82)

In particular, Garrett distinguished between *low* and *high* levels of $L_i$, $T_i$, and $C_i$. For these variables, the 14th percentile in the dataset represented a low value, whereas the 86th percentile represented a high one.[8]

The counterfactual estimates appear in Table 1, which Garrett used to draw three conclusions. First, "government spending was always greater when left-labor power was high than when it was low, irrespective of the level of market integration" (entries in the second row in each table exceeded values in the first row). Second, "the gap between the low and high left-labor power cells was larger in the high trade and capital mobility cases than in the cells with low market integration," implying that "partisan politics had more impact on government spending where countries were highly integrated into the international economy than in more closed contexts." Finally, "where left-labor power was low, government spending decreased if one moved from low to high levels of market integration, but the converse was true at high levels of left-labor power" (1998, 83).

**Table 1 Garrett's Counterfactual Effects on Government Spending (% of GDP)**

| | | Trade | | Capital Mobility | |
|---|---|---|---|---|---|
| | | Low | High | Low | High |
| Left-labor power | Low | 43.1 | 41.9 | 42.8 | 42.3 |
| | High | 43.5 | 44.2 | 43.1 | 44.5 |

Each entry is the expected level of government spending for given configurations left-labor power and trade or capital mobility, holding all other variables constant at their means.

Garrett's counterfactuals go far beyond the customary list of coefficients and *t*-tests, but our tools can help us extract even more information from his model and data. For instance, simulation can reveal whether the differences in values across the cells might have arisen by chance alone. To make this assessment, we reestimated the parameters in Garrett's regression equation[9] and drew 1000 sets of simulated coefficients from their posterior distribution, using the algorithm for simulating the parameters. Then we fixed $L_c$ and $T_c$ at their 14th percentiles, held other variables at their means, and calculated 1000 (counterfactual) expected values, one for each set of simulated coefficients. Following the same procedure, we produced counterfactuals for the other combinations of $L_c$, $T_c$, and $C_c$ represented by the cells of Table 1. Finally, we plotted "density estimates" (which are smooth versions of histograms) of the counterfactuals; these appear in Figure 2. One can think of each density estimate as a pile of simulations distributed over the values government spending. The taller the pile at any given level of government spending, the more simulations took place near that point.

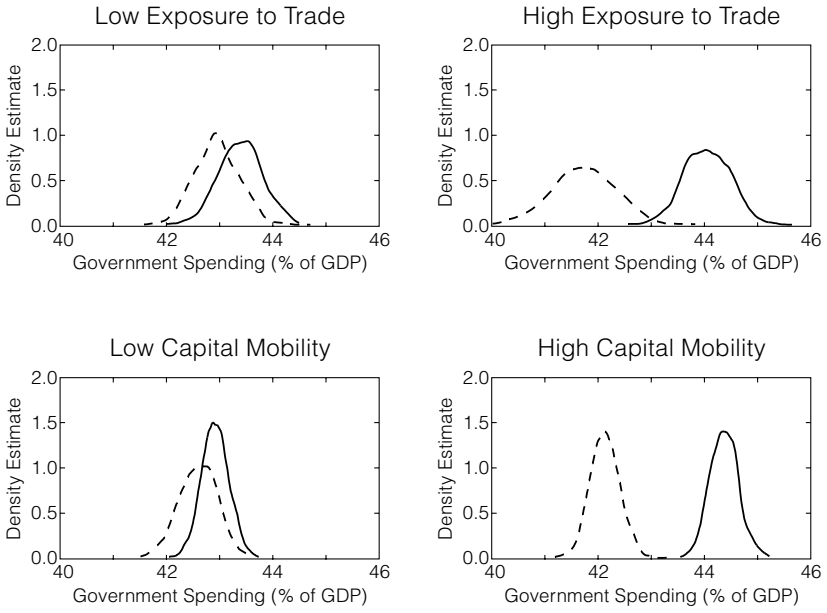
Figure 2 shows that when globalization of trade or capital mobility is low, leftist governments spend only slightly more than rightist ones. More importantly, the

[7]Garrett also focused on two interactions among the variables, $C_iL_i$ and $T_iL_i$, and he included a battery of business cycle and demographic controls, as well as the lagged level of government spending and dummy variables for countries and time.

[8]"So as not to exaggerate the substantive effects" of the relationships he was studying, Garrett "relied on combinations of the 20th and 80th percentile scores" (1998, 82). Unfortunately, due to a minor arithmetic error, the values he reports (1998, 84) correspond only to the 14th and 86th percentiles. To facilitate comparison with Garrett, we use the 14th and 86th percentiles in our simulations.

[9]Our coefficients differed from those in Garrett (1998, 80–81) by only 0.3 percent, on average. Standard errors diverged by 6.8 percent, on average, apparently due to discrepancies in the method of calculating panel-corrected standard errors (Franzese 1996). None of the differences made any substantive difference in the conclusions.

**FIGURE 2 Simulated Levels of Government Spending**




These panels contain density estimates (smooth versions of histograms) of expected government spending for countries where left-labor power is high (the solid curve) and low (the dotted curve). The panels, which add uncertainty estimates to the concepts in Table 1, demonstrate that left-labor power has a distinguishable effect only when exposure to trade or capital mobility is high.

density estimates overlap so thoroughly that it is difficult to distinguish the two spending patterns with much confidence. (Another way to express one aspect of this point is that the means of the two distributions are not statistically distinguishable at conventional levels of significance.) In the era of globalization, by contrast, domestic politics exerts a powerful effect on fiscal policy: leftist governments outspend rightist ones by more than two percent of GDP on average, a difference we can affirm with great certainty, since the density estimates for the two regime-types are far apart. In summary, our simulations cause us to question Garrett's claim that left-labor governments always outspend the right, regardless of the level of market integration: although the tendency may be correct, the results could have arisen from chance alone. The simulations do support Garrett's claim that globalization has intensified the relationship between partisan politics and government spending.

## Multinomial Logit Models

How do citizens in a traditional one-party state vote when they get an opportunity to remove that party from office? Domínguez and McCann (1996) addressed this question by analyzing survey data from the 1988 Mexican presidential election. In that election, voters chose among three presidential candidates: Carlos Salinas (from the ruling PRI), Manuel Clouthier (representing the PAN, a right-wing party), and Cuauhtémoc Cárdenas (head of a leftist coalition). The election was historically significant, because for the first time all three presidential candidates appeared to be highly competitive. Domínguez and McCann used a multinomial logit model to explain why some voters favored one candidate over the others. The following equations summarize the model, in which $Y_i$ and $\pi_i$ are $3 \times 1$ vectors:

$$Y_i \sim \text{Multinomial}(\pi_i)$$

$$\pi_i = \frac{e^{X_i\beta_j}}{\sum_{k=1}^{3} e^{X_i\beta_k}} \text{ where } j = 1, 2, 3 \text{ candidates.} \quad (5)$$

The effect parameters can vary across the candidates, so $\beta_1$, $\beta_2$, and $\beta_3$ are distinct vectors, each with $k \times 1$ elements.[10]

The book focuses on individual voting behavior, as is traditional in survey research, but we used simulation to examine the quantity of interest that motivated

[10]Domínguez and McCann included thirty-one explanatory variables in their model. For a complete listing of the variables and question wording, see Domínguez and McCann (1996, 213–216).

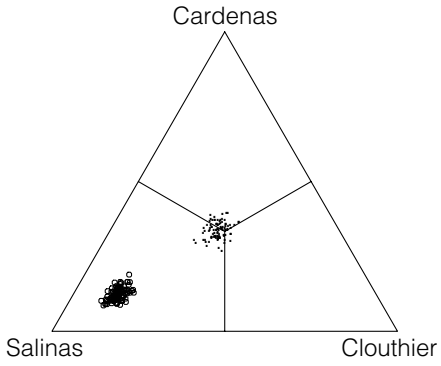
Domínguez and McCann in the first place: the electoral outcome itself. In particular, if every voter thought the PRI was weakening, which candidate would have won the presidency? To answer this question, we coded each voter as thinking that the PRI was weakening and let other characteristics of the voter take on their true values. Then we used the predicted value algorithm to simulate the vote for each person in the sample and used the votes to run a mock election. We repeated this exercise 100 times to generate 100 simulated election outcomes. For comparison, we also coded each voter as thinking the PRI was strengthening and simulated 100 election outcomes conditional on those beliefs.

Figure 3 displays our results. The figure is called a "ternary plot" (see Miller 1977; Katz and King 1999), and coordinates in the figure represent predicted fractions of the vote received by each candidate under a different simulated election outcome. Roughly speaking, the closer a point appears to one of the vertices, the larger the fraction of the vote going to the candidate whose name appears on the vertex. A point near the middle indicates that the simulated election was a dead heat. We also added "win lines" to the figure that divide the ternary diagram into areas that indicate which candidate receives a plurality and thus wins the simulated election (e.g., points that appear in the top third of the triangle are simulated election outcomes where Cárdenas receives a plurality).

In this figure, the o's (all near the bottom left) are simulated outcomes in which everyone thought the PRI was strengthening, while the dots (all near the center) correspond to beliefs that the PRI was weakening. The figure shows that when the country believes the PRI is strengthening, Salinas wins hands down; in fact, he wins every one of the simulated elections. If voters believe the PRI is weakening, however, the 1988 election is a toss-up, with each candidate having an equal chance of victory.

This must be a sobering thought for those seeking to end PRI dominance in Mexico. Hope of defeating the PRI, even under these optimistic conditions, probably requires some kind of compromise between the two opposition parties. The figure also supports the argument that, despite much voter fraud, Salinas probably did win the presidency in 1988. He may have won by a lesser margin than reported, but the figure is strong evidence that he did indeed defeat a divided opposition.[11]

**Figure 3 Simulated Electoral Outcomes**




Coordinates in this ternary diagram are predicted fractions of the vote received by each of the three candidates. Each point is an election outcome drawn randomly from a world in which all voters believe Salinas' PRI party is strengthing (for the "o"'s in the bottom left) or weakening (for the "·"'s in the middle), with other variables held constant at their means.

## Censored Weibull Regression Models

How do wars affect the survival of political leaders? Bueno de Mesquita and Siverson (1995) examine this question by estimating a censored Weibull regression (a form of duration model) on a dataset in which the dependent variable, $Y_i$, measures the number of years that leader $i$ remains in office following the onset of war. For fully observed cases (the leader had left office at the time of the study), the model is

$$Y_i \sim \text{Weibull}(\mu_i, \sigma)$$

$$\mu_i \equiv E(Y_i | X_i) = (e^{X_i\beta})^{-\sigma}\Gamma(1+\sigma) \qquad (6)$$

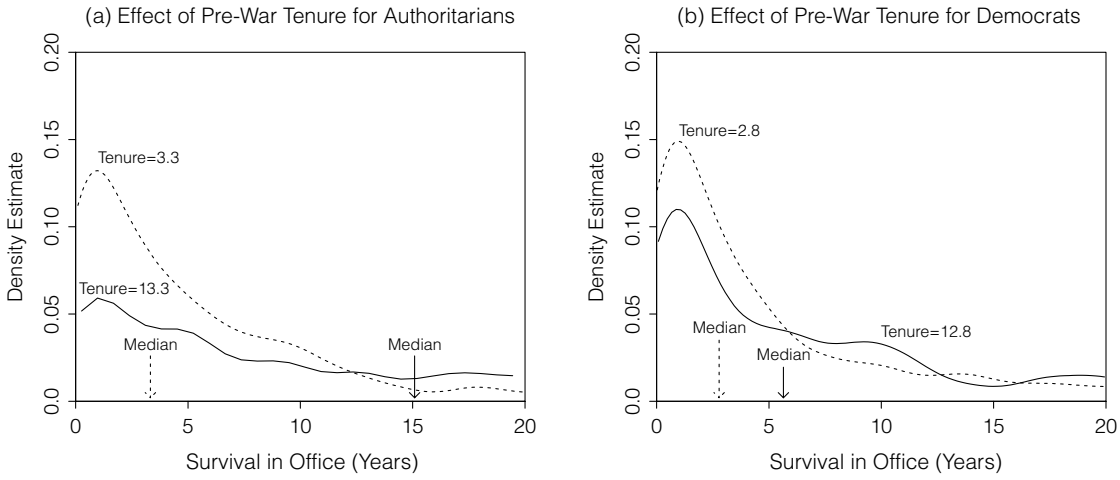
where $\sigma$ is an ancilliary shape parameter and $\Gamma$ is the gamma function, an interpolated factorial that works for continuous values of its argument. The model includes four explanatory variables: the leader's pre-war tenure in years, an interaction between pre-war tenure and democracy, the number of battle deaths per 10,000 inhabitants, and a dummy variable indicating whether the leader won the war.[12] The authors find that leaders who waged foreign wars tended to lose their grip on power at home, but authoritarian leaders with a long pre-war tenure were able to remain in office longer than others.

Bueno de Mesquita and Siverson discuss the marginal impact of their explanatory variables by computing the "hazard rate" associated with each variable. Hazard rates are the traditional method of interpretation in the literature, but understanding them requires considerable statistical knowledge. Simulation can help us calculate more intuitive quantities, such as the number of months that a leader could expect to remain in office following

[11] See Scheve and Tomz (1999) on simulation of counter factual predictions and Sterman and Wittenberg (1999) on simulation of predicted values, both in the context of binary logit models.

[12] The first three variables are expressed in logs.

**FIGURE 4 Regime Type and and Political Survivability in Wars**




Density estimates of the number of years of survival in office for (a) authoritarian and (b) democratic leaders with median pre-war tenure (dotted line) and long pre-war tenure (solid line).

the outbreak of war. As a first step, we predicted the survival time in office for a democrat with the median level of pre-war tenure, holding other variables at their means. After repeating this exercise for an authoritarian leader, we asked what would have happened if the leaders had ten extra years of pre-war tenure under their belts. In each of our four cases, we generated 500 simulations to reflect both estimation and fundamental uncertainty.

The results appear in Figure 4, which displays density estimates of survival time for authoritarians and democrats, conditional on pre-war tenure. The dotted curves correspond to leaders with average levels of pre-war tenure, whereas the solid lines represent densities for leaders with ten extra years of pre-war experience. The arrows in the graphs indicate the median outcome under each scenario. These arrows are further apart in the left panel (a) than in the right one (b), lending strong support to the authors' original claim that prewar tenure matters more for authoritarians than it does for democrats. On average, experienced authoritarians managed to retain power 11.8 years longer than their less experienced counterparts; by contrast, an extra decade of pre-war experience extended the post-war tenure of democrats by only 2.8 years.

Figure 4 also illustrates the value of plotting the entire distribution of a quantity of interest, instead of focusing on a single summary like the mean. Due to assymmetries in the distributions, the modal survival time (the peak of each distribution) does not correspond closely to the median survival time, which is arguably more interesting. The exact nature of the dramatic skewness is also important, since it shows clearly that most survival times are relatively short (under 5 years) and densely clustered, with longer times distributed over a much wider range (5–20 years and more).

## Concluding Remarks

Political scientists have enjoyed increasing success in extracting information from numerical data. Thanks to the work of political methodologists in the last decade or two, we have imported and adapted statistical approaches from other disciplines, created new models from scratch, and applied these models in every empirical subfield. We now collect and analyze quantitative data from a wide range of sources and time periods, and we deposit numerous data sets in archives such as the Inter-University Consortium for Political and Social Research. Most impressively, about half of all articles in political science journals now include some form of statistical analysis, and the methods are becoming increasingly sophisticated and appropriate to the problems at hand.

Unfortunately, our success at developing and implementing new quantitative methods has come at some cost in communication. Many quantitative articles contain impenetrable statistical jargon and unfamiliar mathematical expressions that confuse readers and seem to obscure more of social reality than they reveal. This problem may even account for much of the acrimony between quantitative and qualitative researchers, despite the common goals both groups have in learning about the world. Statistical methods are difficult to learn, harder to use,

and seemingly impossible to present so that nonquantitative social scientists can understand. Few argue against the centrality of statistics for analyzing numerical data, just as few claim any longer that either quantitative or qualitative information will ever be sufficient in isolation. Yet statistical analysts have a responsibility to present their results in ways that are transparent to everyone. In too much research, understanding even the substantive conclusions of sophisticated quantitative models can be challenging at best and impossible at worst.

Political scientists have attacked this communication problem from many angles. Most graduate programs now offer a sequence of courses in political methodology, and an increasing number offer informal math classes during the summer. Methodologists regularly sponsor retraining programs and write pedagogical articles. Yet, all this activity will not make statisticians out of qualitative researchers, nor would it be even remotely desirable to do so.

As a new line of attack, we suggest that the "producers" rather than the "consumers" of statistical research should bear some of the cost of retraining. Our proposals for extracting new information from existing statistical models should enable scholars to interpret and present their results in ways that convey numerically precise estimates of the quantities of substantive interest, include reasonable assessments of uncertainty about those estimates, and require little specialized knowledge to understand.

The methods we propose are more onerous than the methods currently used in political science. They require more computation, and researchers who put them into practice will have to think much harder about which quantities are of interest and how to communicate to a wider audience. But our approach could help bridge the acrimonious and regrettable chasm that often separates quantitative and nonquantitative scholars, and make the fruits of statistical research accessible to all who have a substantive interest in the issue under study. Perhaps most importantly, the proposals discussed here have the potential to improve empirical research and to reveal new facts currently ignored in our already-run statistical procedures. That is, without new assumptions, new statistical models, or new data collection efforts, the interpretive procedures we propose have the potential to generate new conclusions about the political and social world.

## Appendix
## Software

We have written easy-to-use statistical software, called *CLARIFY: Software for Interpreting and Presenting Statistical Results*, to implement our approach. This software, a set of macros for the Stata statistics package, will calculate quantities of interest for the most commonly used statistical models, including linear regression, binary logit, binary probit, ordered logit, ordered probit, multinomial logit, Poisson regression, negative binomial regression, and a growing number of others. The software and detailed documentation are available at http://GKing.Harvard.Edu. We provide a brief description here.

The package includes three macros that are intended to be run in this order:

ESTSIMP estimates a chosen model and generates random draws from the multivariate normal distribution (i.e., computes $\tilde{\gamma}$).

SETX sets $X_c$ to desired values such as means, medians, percentiles, minima, maxima, specified values, and others.

SIMQI computes desired quantities of interest such as predicted values, expected values, and first differences.

These programs come with many options, but to show how easy they can be to use, we provide one brief example. Suppose we have an ordered-probit model in which the dependent variable y takes on the values 1, 2, 3, 4, or 5 and the explanatory variables are x1 and x2. Suppose we want to find the probability that y has the value 4 when x1 = 12.8 and x2 is fixed at its mean, and want a 90 percent confidence interval around that probability. To generate this quantity of interest, we would type the following three commands from the Stata command prompt:

```
estsimp oprobit y x1 x2
setx x1 12.8 x2 mean
simqi, prval(4) level(90)
```

The first line estimates the ordered probit model of y on x1 and x2 and generates and stores simulated values of all estimated parameters. The second line sets x1 to 12.8 and x2 to its mean. The third line computes the desired quantity of interest (a probability value of 4) and the (90-percent) level of confidence associated with an interval to be computed around it.

These programs are very flexible and will compute many more quantities of interest than included in this brief example. The online help gives detailed descriptions. We invite others to write us with contributions to this set of macros to cover additional statistical models or other quantities of interest. We also plan to continue adding to them.

## References

Barnett, Vic. 1982. *Comparative Statistical Inference*. 2nd ed. New York: Wiley.

Blalock, Hubert M. 1967. "Causal Inferences, Closed Populations, and Measures of Association," *American Political Science Review* 61: 130-136.

Bueno de Mesquita, Bruce, and Randolph M. Siverson. 1995. "War and the Survival of Political Leaders: A Comparative Study of Regime Types and Political Accountability." *American Political Science Review* Vol. 89: 841-855.

Cain, Glen G., and Harold W. Watts. 1970. "Problems in Making Policy Inferences from the Coleman Report." *American Sociological Review* 35: 228-242.

Carlin, Bradley P., and Thomas A. Louis. 1996. *Bayes and Empirical Bayes Methods for Data Analysis.* London: Chapman and Hall.

Cowles, Mary Kathryn, and Bradley P. Carlin. 1996. "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review," *Journal of the American Statistical Association*, 91: 883–904.

Domínguez, Jorge I., and James A. McCann. 1996. *Democratizing Mexico: Public Opinion and Electoral Choices.* Baltimore: The Johns Hopkins University Press.

Fair, Ray C. 1980. "Estimating the Expected Predictive Accuracy of Econometric Models," *International Economic Review* 21: 355–378.

Franzese, Jr., Robert J. 1996. "PCSE.G: A Gauss Procedure to Implement Panel-Corrected Standard-Errors in Non-Rectangular Data Sets." Cambridge: Harvard University.

Garrett, Geoffrey. 1998. *Partisan Politics in the Global Economy.* New York: Cambridge University Press.

Kass, Robert E., Bradley P. Carlin, Andrew Gelman, and Radford M. Neal. 1998. "Markov Chain Monte Carlo in Practice: A Roundtable Discussion" The *American Statistician.* 52920: 93–100, 1998 May.

Katz, Jonathan, and Gary King. 1999. "A Statistical Model for Multiparty Electoral Data." *American Political Science Review.* 93:15–32.

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

King, Gary, and Langche Zeng. 1999. "Logistic Regression in Rare Events Data." Unpublished manuscript available at http://GKing.Harvard.edu.

Long, Scott J. 1997. *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks, Calif.: Sage Publications.

Miller, William L. 1977. *Electoral Dynamics in Britain since 1918.* London: Macmillan.

Mooney, Christopher Z. 1996. "Bootstrap Statistical Inference: Examples and Evaluations for Political Science." *American Journal of Political Science* 40: 570-602.

Mooney, Christopher Z., and Robert D. Duval. 1993. *Bootstrapping: A Nonparametric Approach to Statistical Inference.* Newbury Park, Calif.: Sage Publications.

Noreen, Eric. 1989. *Computer-Intensive Methods for Testing Hypotheses.* New York, Wiley.

Rosenstone, Steven J., and John Mark Hansen. 1993. *Mobilization, Participation, and Democracy in America.* New York: MacMillan.

Scheve, Kenneth, and Michael Tomz. 1999. "Electoral Surprise and the Midterm Loss in U.S. Congressional Elections." *British Journal of Political Science* 19: 507–21.

Simon, Julian Lincoln. 1992. *Resampling: The New Statistics.* Arlington, Va.: Resampling Stats.

Simon, Julian Lincoln, David T. Atkinson, and Carolyn Shevokas. 1976. "Probability and Statistics: Experimental Results of a Radically Different Teaching Method." *The American Mathematical Monthly* 83: 733-739.

Sterman, John D., and Jason Wittenberg. 1999 "Path-Dependence, Competition, and Succession in the Dynamics of Scientific Revolution." *Organization Science* 10: 322-341.

Stern, Steven. 1997. "Simulation-Based Estimation." *Journal of Economic Literature* XXXV: 2006–2060.

Tanner, Martin A. 1996. *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions.* 3rd ed. New York: Springer-Verlag.

Tufte, Edward R. 1974. *Data Analysis for Politics and Policy.* Englewood Cliffs, N.J.: Prentice-Hall.

van der Vaart, A. W. 1998. *Asymptotic Statistics.* New York: Cambridge University Press.