How the Techniques You Choose Affect the Answers You Get*

Paper prepared for presentation to the
International Studies Association Conference
February, 1995
Chicago, IL

Kevin A. Clarke
University of Michigan, Department of Political Science
7th floor Haven Hall, Ann Arbor, MI 48109
734 998-1532
kclarke@umich.edu

---

* Apologies to Barbara Geddes.
I would like to thank Lowell Barrington, Bear Braumoeller, Steve Coad, Sue Ellis, Ronie Garcia-Johnson, Ted Hopf, Harvey Schuckman, and in particular, J. David Singer and Kristen Kuntz for their helpful comments and advice.

Hardly anyone takes data analyses seriously. Or perhaps more accurately, hardly anyone takes anyone else's data analyses seriously. (Leamer, 1983)

The Problem

Issue areas in political science change slowly. Contemporary scholars are often interested in the same questions that were asked by their 1950s counterparts. Americanists are still interested in elections, comparativists in Western European political parties, and world politics scholars in war. What changes more rapidly, however, are the methods and techniques used to study these questions. Econometrics is a case in point. Econometric theory and practice have changed dramatically over the last fifty years.

Given that questions change slowly while econometric practices do not, we are bound to see different econometric techniques being applied to similar questions with similar data. The result is often confusion. Different techniques applied to similar data (or even the same data) frequently yield different results.[1] We might assume, as some researchers do, that newer techniques are simply 'more appropriate' to the question being asked than older techniques. Given the rate of change in econometric practice, these 'more appropriate' techniques were probably not available to the early researchers in the field. Those early scholars used the 'most appropriate' techniques available in their time. The confusion of inferences and conclusions is therefore the consequence of changing econometric practice. If this description is accurate, we can have little confidence in the results of quantitative statistical analyses unless we posit an 'end-to-statistical-history', that is, a point beyond which the development of statistics is unnecessary and impossible.

[1] By different results, I do not mean simply different numbers. An OLS regression and a Poisson regression may yield different coefficients that when interpreted mean the same thing. By different results, I mean coefficients that produce different *inferences*.

This lack of confidence is most troubling when we consider the impact of quantitative studies on the policy arena. This is particularly true of world politics where noisy data and measurement error already undermine the policy relevance of most empirical work (King, 1989b). Some may argue that these problems along with those posed by specification error are more damaging to policy relevance than different conclusions drawn from changing econometric techniques. The former problems, however, while remaining unsolved, are at least fairly well understood. The use of specification tests, robustness checks, and measurement error models are now widespread. The effect of changing econometric practices, though, is not as well understood and in some cases, not even acknowledged.

This ambiguity of our collective results inhibits the progressivity of our field. If we cannot take the results of other scholars' data analyses seriously, then we cannot build upon their findings and we end up reinventing the wheel over and over again. Consistent results and inferences are the basis of a cumulative science; without them our field must remain in its infancy.

This paper addresses the impact of changing econometric theory and practice on the inferences drawn from quantitative studies. I demonstrate this impact using an example drawn from the literature of world politics. The next section presents the argument, followed by the data analysis, counter-arguments, and partial solutions.

The Argument

Philosophers of science, if not scientists themselves, have long understood the dependence of observation upon theory (Chalmers, 1982). That is, any observation made by an observer depends in part upon his "past experience, his knowledge and his expectations" (Chalmers, 1982). The classic discussion of this idea in N.R. Hanson's

Patterns of Discovery (1961) begins with the question, “Do Kepler and Tycho [Brahe] see the same thing in the east at dawn?” Hanson argues that the answer is no. While Kepler and Tycho may view the same phenomenon, seeing is an experience while a retinal reaction is simply a physical state. “People, not their eyes, see” (Hanson, 1961). Because seeing occurs instantaneously, this difference is more than a difference in interpretation, it is rather a difference in conceptual organization or context. While the elements of Kepler’s and Tycho’s experiences are identical, their conceptual organization is vastly different (Hanson, 1961). Kepler and Tycho do see different things in the east at dawn.

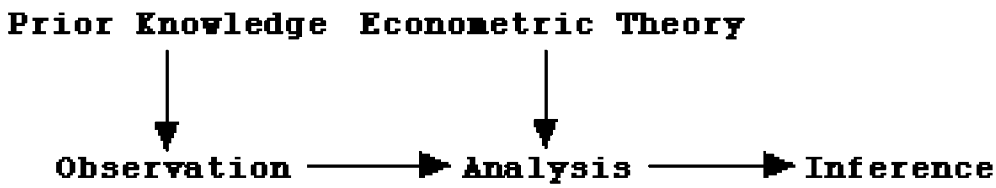
Observation then is as fallible as the theory it presupposes and therefore not a secure basis for science. A concrete example may make this point clearer. The statement “state X in alliance Y initiated a war against state Z” presupposes quite a bit of low-level theory. It assumes that political entities called states exist, that states can join political entities called alliances (which must themselves exist), and that states may initiate a social phenomenon called war. Remember, however, that political entities called states (countries) did not even exist until the Peace of Westphalia.

Observation, therefore, presupposes theory and as noted earlier, this concept is well understood by philosophers of science and scientists alike (Boyd, 1993).[2] There are cases, however, where more formal theory intervenes between theory-guided observations and the conclusions or inferences drawn from those observations. If this theory or framework undergoes significant changes or is simply wrong, then the conclusions based on that framework will either significantly change or be wrong. For instance, computerized models allow meteorologists to turn weather observations collected at time *t* into forecasts for time *t+1*. If these models change or are incorrect, the resulting forecasts will either change or be wrong. A diagram will help illustrate this point. In

[2] In a sense, since all conceptions of science are plagued by this critique, the problem drops out.

Diagram 1, prior knowledge and experience, as noted earlier, affect the process of observation. After theory-guided observation comes analysis which is performed through the framework of formal theory (econometric theory, in this case). It is this interaction between formal theory and analysis and its effect upon inference that is the concern of this paper.

Diagram 1. ***The Inference-generating Process***



Like weather-forecasting models, econometric analysis is an example of a formal theory intervening between observation and inference.

In quantitative research, econometric theory stands between the observations we gather from the social world around us and the inferences we make about that social world. That is, when faced with large amounts of data, we use statistics to summarize and describe accurately the information contained in those data (Achen, 1982). This information, in turn, allows us to make inferences regarding the causal patterns contained in the data. As econometric theory and practice change, we are likely to get a different description of our data, which is likely to lead us to different conclusions regarding our question of interest. In fact, if we are strict falsificationists, statistics may induce us to throw out a theory which later statistics would induce us to keep. The opposite is also a plausible, and even probable, scenario. We simply cannot know for sure either way.

Ideally, social scientists would use statistics as an inference-making tool. Statistics would, like a hammer, always work in the same way. While there are large hammers and small hammers, some more appropriate for some jobs than others, most simply deliver a concentrated force upon a specific target with an expected result. Although hammers are no doubt easier and cheaper to produce than they once were, there have been few innovations in either hammer design or hammering technique. The same cannot be said for statistics which has seen dramatic change in the last few decades.

Kotz and Johnson (1992) write of "a growing, and already embarrassingly noticeable, diversification among the statistical sciences which borders on fragmentation." They also write that the character of statistics has "changed dramatically over the last 100 years and at an accelerated rate during the last 25 years." A comparative glance between any econometrics text from the 1960s and any comparable econometrics text from the 1980s will bear out Kotz and Johnson's claims. In the preface to their second edition, Judge *et al* (1985) write, "the pace of developments in econometrics over the last *four* years have made it necessary to rewrite and restructure the first edition completely."[3] Additions to Judge *et al* (1985) include Bayesian inference, time series, simultaneous equation statistical models, distributed lags, and nonlinear estimation. While it would be exceedingly difficult to gauge the pace of change in econometrics, it is probably safe to say that this "vigorous development" has slowed little in the last few years and shows few signs of slowing down anytime in the near future.

The point here is that statistical inference is based on a theory that unlike hammer-theory, changes constantly. This process quite naturally affects our results.[4] The first step in understanding this process is to understand at what points changing statistical

[3] Italics added.
[4] I illustrate this process in a later section.

methodology affects our results. There are two such points of interest: choice of statistical model and the interpretation of results. As new econometric models are developed, they eventually replace or augment the older models in the statistical arsenals of most scholars. Quite simply, if these new models did not produce different and hopefully superior results, there would be no point in switching to them. One such case should be familiar to most researchers. At one time scholars faced with a dichotomous dependent variable would have relied on either simple OLS regression or the linear probability model; contemporary scholars usually opt for the now familiar logit and probit models (Greene, 1993). The results produced by the logit and probit models are different from those produced by OLS or linear probability.

Changing methodology also affects our interpretation of results. Statistics such as the correlation coefficient, R^2, or standardized beta, have seen their interpretations shift from absolute measures of causal strength to sample-specific ones. Conclusions based on these statistics have therefore changed as well. At one time, the best regression model was the one with the largest R^2. Today, “the best regression model usually has an R^2 that is lower than could be obtained otherwise” (King, 1986).

These two categories (statistical models and interpretation), however, are not mutually exclusive. Changes in statistical interpretation lead to changes in statistical modeling. Stepwise regression is one such example. Designed to maximize R^2, stepwise regression fell from the statistical canon as the interpretation of R^2 changed from percent of variance explained to the geometric shape of the regression points. The opposite may also be true. The use of the atheoretical stepwise procedure may have undermined confidence in the R^2 statistic.

Changes in statistical theory affect the models we choose and the interpretations we give to our results.

The Analysis

I will present, in this section, an illustration of the process described above. Taking a data-based article from the mid-1960s, I will reestimate the model described therein using two econometric techniques that were, for the most part, unavailable to the original authors. Using the same data and model specification, I can attribute any divergent inferences to the effect of changing estimation practices.

The question of how much change constitutes a different result should be addressed here. To make the test easy, I require only that the *pattern* of results (the direction of the coefficients and their significance or nonsignificance) remain consistent. A more difficult test would require coefficients that remain consistent in direction, significance, and impact. That is, a coefficient would not only have to remain positive or negative, it would also have to produce the same change in the dependent variable. If the analysis presented here fails to pass the first test, it has certainly failed the second.
This example does not, however, constitute a rigorous test of the process in question. Some of the problems associated with single case studies - little logical constraint for theory-building and little inferential constraint for theory-testing - apply here (Achen and Snidal, 1989). Rather, the analysis simply illustrates the dangers of drawing quantitative inferences with changing econometric methods.

The Design

The analysis to be reestimated here is the classic Singer and Small article of 1968, “Alliance Aggregation and the Onset of War, 1815-1945.” This particular study was chosen for three reasons. First, most of the original data used in the study is readily available. Since Singer and Small began collecting quantitative world politics data, the database has been expanded and updated directly and continuously. That is, once a study was finished, the data used for that study was not locked away intact; new data were added directly to it (including revisions of old data). The consequence of this method of data storage is that the data used in any of the older studies are inseparable from the updated data. Retrieving such data is next to impossible. In the mid-1960s, however, journals were still publishing articles comprised simply of collected data. Fortunately, the data used in the 1968 “Alliance Aggregation” article were published in this way and it is from these articles that the data are drawn (Singer and Small, 1966). Second, a glance at the Social Science Citation Index shows that the “Alliance Aggregation” article has had a profound impact on the alliance studies that followed it. Third, the study uses, by today’s standards, an almost primitive data analysis technique - bivariate correlations. This technique, however, was not only standard for its time, it may have been ahead of its time for world politics.

The first step in this demonstration is to replicate the original article in order to ensure that the data I use are substantively equivalent to the data used in that study. The next step will be to reestimate the model using different econometric methods. The first of these methods will be linear regression; at one time, a very common method of estimation in world politics (King, 1989b). The second method will be a generalized event count that was only recently developed to handle noncategorical discrete variables such as the number of wars per year (King, 1989b). It is essential to note that each of these techniques were or are considered to be “appropriate” for the question of alliances and war.[5]

---

[5] Whether or not a particular technique is considered “appropriate” for a question depends on a complex interaction between statisticians, political science statisticians, and political science consumers.

The Replication

Despite the fact that the raw data were published in a separate article, replicating the original study was not easy. Two of the variables used in the Singer-Small article, *bipolarity initial* and *bipolarity alternate*, were never used before or after the 1968 article and were never published in raw form. The variables are constructed by identifying the targets of defensive alliances among major powers. There was enough ambiguity in the original data (48 out of 130 years), however, to necessitate the construction of two indicators, initial and alternate. Therefore, even if the targets could be accurately identified again, there would be no way of knowing which targets ended up in which variable.[6] These variables are for all intents and purposes unreplicable.

The time lags and averages used in the Singer-Small study also present a problem. Data were collected for the period from 1815-1939 with the exception of World War I (1915-1919). The dependent variables used by Singer and Small are three-year cumulative time lags. That is, the dependent variables for 1851 reflect "the amount of war which began between the beginning of 1852 and the end of 1854" (Singer-Small, 1968). This calculation, in and of itself, presents no problem. The ambiguity enters when attempting to decipher what figures were used for 1937, 1938, and 1939 where data for the following three years are not available.[7] To resolve the dilemma, I assume that the dependent variables for 1937 reflect the amount of war begun between the beginning of 1938 and the end of 1939, that the dependent variables for 1938 reflect the amount of war begun between the beginning and end of 1939, and the same for 1939. I assume the same for the 1912-1914 period.

---

[6] As I discuss later, the inability to replicate these variables is not a crushing blow to this analysis.

[7] As data were not collected for either of the world wars, this same problem exists for 1912, 1913, and 1914.

The independent variables used by Singer and Small are three-year averages. That is, for 1908, the independent variables are an average of 1906, 1907, and 1908. Again, the problem here is not the actual calculation but rather deciphering what was done with 1815 and 1939 where data for 1814 and 1940 were not available.[8] Here I assume that 1815 is an average of 1815 and 1816 while 1939 is an average of 1938 and 1939. I assume the same for 1914 and 1920.

Despite these problems, it was possible to replicate all five dependent variables and five out of seven independent variables. When the analysis is rerun, only 62% of the replicated coefficients (118 out of 190) are within 10% of their original values. Most of the large discrepancies, however, occur for estimated coefficients that are close to zero. For instance, the original correlation coefficient between *nation-months of war begun - all* and *percent of major with minor* for the total system between 1900-1939 is 0.05. The replicated coefficient is 0.03. While this is a 40% change in absolute value, there is no substantive difference between the two coefficients.
In addition, when each set of replicated coefficients is correlated with the corresponding set of original coefficients, nine out of ten sets correlate at 0.95 or higher.[9] The tenth set correlates at the 0.88 level. There are no sign changes. I can therefore conclude that the replicated data are both substantively, and in most cases, absolutely, similar to the original data. Differences in inferences then can be attributed to different methods rather than differences in data.

The Bivariate Correlation Results

In "Alliance Aggregation and the Onset of War, 1815-1945" Singer and Small test two hypotheses: *the greater the number of alliance commitments in the system, the more war*

---

[8] Again the same question goes for 1914 and 1920.
[9] Four correlate at 1.0 and one each at 0.99, 0.98, 0.97, 0.96 and 0.95.

the system will experience and *the closer to pure bipolarity the system is, the more war it will experience*. I consider only the first hypothesis as the replication problem prevents me from looking at the second hypothesis. This focus will not be a problem as Singer and Small treat the two hypotheses and their corresponding indicators as distinctly separate. Singer and Small test the alliance hypothesis by correlating five dependent variables with five independent variables for three time periods and two spatial domains. The dependent variables are: *the number of nation-months of war begun for all nations, the number of nation-months of war begun for major nations, the number of battle deaths for all nations, the number of battle deaths for major nations*, and *the number of wars*.[10] The first four dependent variables measure the magnitude and severity of war while the last variable measures the frequency of war.[11]

The independent variables in the study include: *the percentage of all nations having at least one alliance of any class with any type of nation, major or minor, the percentage of all nations having at least one defensive pact with any type of nation, the percentage of major powers having at least one alliance of any class with another major power, the percentage of major powers having at least one defensive pact with another major power, the percentage of major powers having at least one alliance of any class with any minor power*.[12]

These twelve variables were correlated for the period from 1815 to 1939, the period from 1815 to 1899, and the period from 1900 to 1939. The two spatial domains are the total system and the central system, which does not include peripheral members of the nation-state system. I will only report results for the total system for the 1900-1939 period

---

[10] Each dependent variable is constructed as a three-year cumulative time lag.

[11] Those interested in how these indicators were constructed should consult the Correlates of War Data Handbook.

[12] Each independent variable is constructed as a three-year average.

because it is in this time period that Singer and Small find the results most strongly in favor of their hypothesis that alliances affect the onset of war. The results from Singer-Small Table 10 are reproduced below in Table 1.[13]

Table 1. ***Total System, 1900-1939: Correlations between Alliance Indicators and Magnitude, Severity, and Frequency of War (N=35)***

	% of all in any Alliance	% of all in Defense Pact	% of Majors in any Alliance	% of Majors in Defense Pact	% of Majors with Minor
Nation-Months War - All	*53*	43	24	05	43
Nation-Months War - Majors	*46*	*48*	35	16	*47*
Battle Deaths for All	*56*	*48*	29	08	*46*
Battle Deaths for Majors	*51*	*48*	31	13	45
Number of Wars	18	29	*50*	26	*54*

The replicated results for *nation-months of war begun - all* and *the number of wars* appear in Table 2.

The replicated results closely match the originals. One additional variable, *percent of all in a defense pact*, when correlated with *nation-months of war begun - all* reaches conventional significance where it just missed in the earlier study.

[13] The italics indicate those variables which equal or exceed the 0.01 level requirement. (These italics are from the original.)

Table 2. ***Total System, 1900-1939: Replicated Correlations between Alliance Indicators and Magnitude, Severity, and Frequency of War (N=35)***

	% of all in any Alliance	% of all in Defense Pact	% of Majors in any Alliance	% of Majors in Defense Pact	% of Majors with Minor
Nation-Months War - All	*55*	*45*	25	03	42
Number of Wars	23	35	*58*	37	*61*

It is best to let Singer and Small interpret these coefficients in their own words for, as noted above, changing econometric method and practice have a significant effect upon interpretation. Since the replicated coefficients are extremely close to the coefficients reported in the original study, the passages quoted below apply equally well to both sets of results. Singer and Small write:

> ...the percentage of all nations in any class of alliance, as well as in defense pacts, correlates highly with all four magnitude and severity measures in every case but one (.43 with nation-months of war for all). Likewise, the major-minor figure predicts well to three of these four dependent variables.

The authors conclude that:

> ...the evidence seems to be relatively unambiguous...if we look at the twentieth century segment only, the hypothesis is rather strongly confirmed...we find that the percentage of all nations in the system having at least one alliance with any other nation predicts to the amount of war on all [four] of the possible occasions. And the percentage of major powers having at least one alliance with a minor does likewise on [three] out of [four] possible occasions. Combining this powerful tendency with the fact that there are quite a few more correlations that are only slightly weaker and that not a single negative correlation appears,

> we may only conclude that the well-accepted hypothesis has indeed been borne out by our historical evidence.[14]

The authors are therefore confident that, using an appropriate statistical technique, they have demonstrated that alliances have a significant positive effect on the onset of war.

The OLS Reestimation

When ordinary least-squares regression was introduced to world politics, it constituted an enormous improvement over previously used methods (such as bivariate correlations).[15] A discussion of these improvements is appropriate at this juncture. The simple correlation coefficient is the covariance between two variables divided by their respective standard deviations:

$$\mathrm{r}_{yx} = \frac{\mathrm{cov}(X,Y)}{S_x S_y}$$

The value of the correlation coefficient will be larger in absolute value not only when the relationship between X and Y is stronger in absolute value, but also when the variance of the other systematic and random variables that affect Y is smaller and when the variance of X is larger (Hanushek and Jackson, 1977). The size of the correlation coefficient therefore depends in important ways upon the variances of the included and excluded variables.

As the development of the least squares estimator should be familiar to most readers, I do not pursue that development here.[16] Suffice it so say that OLS regression is superior to the simple correlation coefficient because it brought statistical control, interpretability,

[14] Singer and Small are referring here to only the indicators of magnitude and severity. They are also referring, in the original, to both the total and central system results. I have changed the numbers so that they refer to only the total system.

[15] OLS regression certainly existed in the 1960s but it had yet to be utilized by this particular branch of political science.

[16] Interested readers should refer to Hanushek and Jackson (1977).

and robustness to quantitative studies in world politics. In addition, while the correlation coefficient measures only the degree of association between two variables, OLS measures the magnitude or strength of the relationship (Hanushek and Jackson, 1977). It is essential to remember, however, that both the correlation coefficient and OLS regression assume that the underlying relationship is linear. For a time, OLS regression was considered to be the appropriate technique for most world politics questions (King, 1989b).

Although in moving from bivariate correlations to OLS regression, I am moving from bivariate estimation to multivariate estimation, the usual problems of model specification are absent here. Singer and Small have, in effect, spelled out the relevant variables; they simply lacked the technique to estimate the effects of these variables all at once. The independent variables for my model, therefore, will include: *the percentage of all nations having at least one alliance of any class with any type of nation, major or minor*; *the percentage of all nations having at least one defensive pact with any type of nation*; *the percentage of major powers having at least one alliance of any class with another major power*; *the percentage of major powers having at least one defensive pact with another major power*; and *the percentage of major powers having at least one alliance of any class with any minor power*.[17]

In the original bivariate study, these independent variables are correlated with all five dependent variables. In the interest of space and time, I consider only two of these dependent variables. The first will be *nation-months of war begun - all*. As this variable correlates at the 0.95 level or above with the other three measures of magnitude and severity of war, little information will be lost by not estimating equations with those variables. The second equation will use *the number of wars begun* as the regressor. I

[17] These percentage variables are scaled from 1 to 100 as opposed to 0.01 to 1.0.

examine this variable because it only correlates at about the 0.30 level with the other dependent variables. I suspect this equation then will generate different results from the first equation.

The results of the first equation are in Table 3 and the results of the second equation are in Table 4.

Table 3. ***Total System, 1900-1939: OLS Estimates of Alliance Indicators on the Number of Nation-Months of War Begun (N=35)***[18]

	OLS Estimates	Heteroskedastic-Consistent Standard Errors
% of all in any Alliance	*13.380*	*6.425*
% of all in Defense Pact	-18.910	30.620
% of Majors in any Alliance	-2.843	6.159
% of Majors in Defense Pact	2.737	3.917
% of Majors with Minor	6.677	11.094
Constant	-147.997	213.705

Interpretations of these results is straightforward. For a one unit change in *percent of all in any alliance*, there is a 13.4 increase in the number of nation-months of war begun within a three year period. Judging from the relatively low standard error, this effect seems to be fairly systematic in these data. For a one percentage point change in *percent*

[18] Conventionally significant results are italicized.

of all in a defense pact, there is a -18.9 decrease in the number of nation-months of war begun within a three year period. The standard error for this variable, however, permits no confidence that this effect is systematic in these data. The three remaining regression coefficients are neither large nor conventionally significant.

Comparing these results to the bivariate results in the first row in Table 2, we can see that in the bivariate case all five coefficients are positive and three out of four are either significant or close to it.[19] The same cannot be said for the OLS results. In the reestimation, only *percent of all in any alliance* is both statistically significant and in the hypothesized direction. The coefficient on *percent of all in a defense pact* is large, but it is not conventionally significant and it is not in the hypothesized direction.

Alternative explanations for these results such as multicollinearity, which could account for the large standard errors, and highly correlated measurement error in proxy variables, which could account for the sign change, can be rejected. Examination of the R^2_{aux} statistic for the auxiliary regressions of each independent variable on all the other independent variables reveals that collinearity is not a concern. That is, the R^2_{aux} statistics are less than the overall R^2 of the regression (Greene, 1993). By the same token, while *percent of all in any alliance* and *percent of all in a defense pact* are correlated at the 0.70 level, there is no reason to assume that these variables have been measured with significant error, that those errors are highly correlated, or that *percent of all in a defense pact* is the less reliable indicator, which are the conditions under which a sign change can be expected (Achen, 1985). Furthermore, even if we posit highly correlated error, the coefficient that flips should not only be negative, it should be smaller as well, leaving a net positive effect - assuming the true effect is positive (Achen, 1985). The net effect here is negative.

[19] The cutoff point is 0.45.

Similar arguments may be made when using *the number of wars begun* as the dependent variable. Comparing these results (reported in Table 4) to the bivariate results in the second row in Table 2, all five

Table 4. ***Total System, 1900-1939: OLS Estimates of Alliance Indicators on the Frequency of War Begun (N=35)***

	OLS Estimates	Heteroskedastic-Consistent Standard Errors
% of all in any Alliance	-0.011	0.018
% of all in Defense Pact	-0.089	0.086
% of Majors in any Alliance	0.016	0.017
% of Majors in Defense Pact	-0.010	0.011
% of Majors with Minor	0.053	0.031
Constant	0.073	0.600

coefficients in the bivariate case are positive and two out of three are large and conventionally significant. The results from this estimation show only two out of the five coefficients are in the hypothesized direction, none of the coefficients is large, and only one coefficient, that on *percent of major with minor*, approaches statistical significance. The collinearity diagnostics remain the same as do the arguments regarding measurement error.

Presented with the results of the bivariate correlation analysis in Table 2, one would infer, as did Singer and Small, that *percent of all in any alliance*, *percent of all in a defense pact*, and *percent of major with minor* have an important effect on *nation-months of war begun - all*. One would also infer that *percent of majors in any alliance* and *percent of major with minor* have an important effect on *the number of wars begun*. Presented with the OLS results in Tables 3 and 4, one would infer that only *percent of all in any alliance* had an important effect on *nation-months of war begun - all* and that only *percent of major with minor* has an effect on *the number of wars begun* but that effect is only somewhat systematic. By the criteria noted earlier, I can conclude that the bivariate analysis and the OLS regression have produced different conclusions.

The MLE Reestimation

The wide dissemination of event count models had to wait until the mid-1980s. The fact that neither Judge *et al* (1985) nor the second edition of Kmenta's Elements of Econometrics (1986) provides a discussion of event count models is indicative of the model's limited application prior to this period. Given the model's relative obscurity, I will provide a brief overview of its assumptions and development as well as the reasons for preferring it to OLS in this case.[20]

The event count model is considered more appropriate for the question of alliances and war because it is designed to handle noncategorical discrete dependent variables such as the number of wars begun or the number of nation-months of war begun as opposed to OLS which is designed to handle continuous, interval level dependent variables (King, 1989b). The problem stems from the fact that the dependent variables used in the Singer-Small study (*nation-months of war begun* and *the number of wars begun*) are bounded by zero at the low end and are unbounded at the high end. That is, in any given time period,

[20] A more detailed development can be found in King (1989). This analysis draws on King's discussion.

there can be no war or some war but there cannot be less than no war. Using OLS to estimate the effect of alliances on these dependent variables would produce nonsensical results such as negative levels of war. In addition, using OLS to estimate models with bounded dependent variables violates the assumption of homoskedasticity (error terms with equal variance) required for OLS. The results are inefficient estimators and biased standard errors. The use of event count models avoids these problems and brings with them the desirable properties of maximum likelihood estimators: consistency, asymptotic efficiency, and asymptotic normality (Hanushek and Jackson, 1977).[21]

For event count models, the form of uncertainty in the random variable Y (the dependent variable) across repeated experiments or the stochastic component of the model is the Poisson distribution. Three assumptions must be met for an observation-generating process to be a Poisson distribution (King, 1989a). First, during each observation period *i*, the rate of event occurrence λ_i remains constant. That is, there is no contagion between the occurrence of events. Second, the probability of two events occurring at precisely the same instant is very close to zero. Third, zero events have occurred at the start of the period. For *n* observations, the Poisson distribution is written as:

$$\Pr(Y \mid \lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \qquad (1)$$

The statement of how the parameter of interest varies over observations as a function of the independent variables or the systematic component of the model is:

$$E(Y_i) \equiv \lambda_i = \exp\ (x_i \beta) \qquad (2)$$

[21] Using the event count model does, however, create other problems. Due to the manner in which the dependent variables are constructed, autocorrelation is built into the analysis. The OLS coefficient estimates are unbiased under these conditions. There is no such guarantee for the MLE estimates. We therefore face a trade-off between robustness and appropriate assumptions.

The technical reason for choosing this function is to constrain the expected number of events to be positive.[22] The substantive reason for this function is "so that a fixed change in X_j would have a greater effect on the expected value if the expected value were larger" (King, 1989a). The instantaneous effect of X_j on Y is therefore:

$$\frac{\delta\lambda_i}{\delta x_i} = \exp(x_i\beta) * \beta = \lambda_i\beta \qquad (3)$$

Substituting equation (2) into equation (1), we get the likelihood function:

$$L(\tilde{\beta} \mid y) = \prod_{i=1}^{n} \frac{e^{-\exp(x_i\tilde{\beta})}\exp(x_i\tilde{\beta})y_i}{y_i!} \qquad (4)$$

We take the log of equation (4) for computational ease and get equation (5) which can be maximized with respect to β with numerical methods:

$$\ln L(\tilde{\beta} \mid y) = \sum_{i=1}^{n} \left\{ (x_i\tilde{\beta})y_i - \exp(x_i\tilde{\beta}) \right\}^{23} \qquad (5)$$

This model was estimated with both *nation-months of war begun - all* and *the number of wars begun* serving as the dependent variable. The independent variables remain the same as in the previous estimation. The results of the event count model using *nation-months of war begun - all* as dependent variable are presented in Table 5.

Interpretation of these coefficients is relatively more involved that OLS, but still easily interpretable. The instantaneous effect of X_j on Y is given in equation (3).

---

[22] The exponential is the only function that is its own derivative.
[23] -ln y_i! can be dropped because it is not a function of ß.

Table 5. ***Total System, 1900-1939: Poisson Regression of the Number of Nation-Months of War Begun on Alliance Indicators (N=35)***

	MLE Estimates	Heteroskedastic-Consistent Standard Errors
% of all in any Alliance	-0.0013	0.0433
% of all in Defense Pact	0.0251	0.2270
% of Majors in any Alliance	-0.0197	0.0294
% of Majors in Defense Pact	-0.0010	0.0166
% of Major with Minor	0.1113	0.0585
Constant	-0.6759	2.9515
log-likelihood	26987.9480	

Therefore, by multiplying the coefficient on *percent of all in any alliance* (-0.0013) by the mean of the dependent variable (168) by a ten percentage point change in the percentage of nations involved in at least one alliance produces a 2.2 decrease in *the number of nation-months of war begun*. A similar change in the percentage of nations involved in at least one defense pact produces a 42.2 increase in *the number of nation-months of war begun*. The other coefficients may be interpreted in a similar manner. These results certainly do not agree with those produced by the other techniques. In the OLS regression, *percent of all in any alliance* had a relatively large positive effect on the dependent variable. Similarly, the *percent of all in a defense pact* had a large negative effect on the dependent variable in the OLS regression, which has been contradicted by the maximum likelihood estimation.

Consider the whole pattern. The Singer-Small hypothesis posits that the greater the number of alliance commitments in the system, the more war the system will experience. This hypothesis is strongly confirmed in the bivariate analysis where all five coefficients are positive and three out of five coefficients hit or approach conventional significance (see the first row in Table 2). In the OLS regression analysis, three out of five coefficients are positive and only one variable, *percent of all in any alliance,* is conventionally significant (see Table 3). Only two out of five coefficients are positive for the MLE estimation. The coefficient on only one variable, *percent of major with minor*, approaches conventional significance in the event count model. The conclusion to be drawn here is that each of these models would produce different inferences regarding the effects of alliances on the onset of war.

The results from the model using *the number of wars begun* as dependent variable are presented in Table 6.

For this equation, a ten percentage point change in the *percent of all in any alliance* produces -0.0175 x 1.2 (the mean of the dependent variable) x 10 = 0.21 fewer wars. A ten percentage point change in the *percent of all in a defense pact* produces -0.0479 x 1.2 x 10 = .58 fewer wars. The other variables are interpreted in a similar manner. The pattern of these results is fairly close to the pattern produced by the OLS regression but again, with the exception of *percent of major with minor,* none of the coefficients is large and none is conventionally significant.

Table 6. ***Total System, 1900-1939: Poisson Regression of the Frequency of War Begun on Alliance Indicators (N=35)***

	MLE Estimates	Heteroskedastic-Consistent Standard Errors
% of all in any Alliance	-0.0175	0.0177
% of all in Defense Pact	-0.0479	0.0840
% of Majors in any Alliance	0.0138	0.0164
% of Majors in Defense Pact	-0.0104	0.0091
% of Major with Minor	0.0572	0.0309
Constant	-1.8752	0.7417
log-likelihood	-27.4267	

It could be argued that the event count model presented here is too narrowly specified and that the data are overdispersed (King, 1989b). That is, the dependent variable in this case, *nation-months of war begun - all*, may deviate from the assumption of independence indicating that the occurrence of some nation-months of war makes others more likely. The model to handle such overdispersion is a generalized event count that incorporates a scalar dispersion parameter.[24] King (1989a), however, lists some caveats for those scholars using this model. First, by estimating another parameter, we are placing more demands upon the data thereby decreasing the available degrees of freedom even further

[24] This same model may also be used to handle underdispersion.

(to 28 in this case). Second, one cannot automatically assume that the desirable properties of MLE apply without further analysis. King (1989a) does note, however, that consistency can still be proved given these nonstandard conditions.

With these warnings in mind, the results from the generalized event count model with *nation-months of war begun - all* as dependent variable are presented in Table 7.

Table 7. ***Total System, 1900-1939: Generalized Poisson Regression of the Number of Nation-Months of War Begun on Alliance Indicators (N=35)***

	MLE Estimates	Heteroskedastic-Consistent Standard Errors
% of all in any Alliance	-0.0208	0.0382
% of all in Defense Pact	-0.0683	0.1583
% of Majors in any Alliance	-0.0052	0.0263
% of Majors in Defense Pact	-0.0153	0.0138
% of Major with Minor	0.1101	0.0520
σ^2	5.9783	0.2674
Constant	1.8818	1.1991
log-likelihood	30684.7691	

Using the generalized event count model was apparently a good idea judging from the result that $\square^2$ is significantly greater than 1. The GEC model, however, has produced yet another set of alternative findings. The only large, positive variable in the event count model, *percent of major with minor*, has doubled in size and is conventionally significant. The coefficient on the *percent of majors in any alliance* has changed signs. None of the other coefficients is either large or significant.

Summary

The result of these analyses is obviously confusion. The findings from the three estimations using *the number of wars begun* as dependent variable are presented in Table 8 and the findings from the four estimations using *the number of nation-months of war begun - all* as dependent variable are presented in Table 9.

The only inference we can make with any kind of certainty is that the *percent of major with minor* appears to have some positive effect on *the frequency of war begun*. What this independent variable actually measures, however, is unclear and the theoretical reason for its inclusion was left largely unexamined in the Singer and Small study. As noted earlier, the pattern of the effects of the other independent variables on *the frequency of wars begun* is similar in the OLS and MLE estimations, neither of which concur with the bivariate correlations.

Table 8. ***Results of Various Estimation Procedures for the Number of Wars Begun Regressed on Alliance Indicators (coefficient direction, conventional significance)***

	Bivariate Correlations	Ordinary Least Squares	Event Count
% of all in any Alliance	+, no	-, no	-, no
% of all in Defense Pact	+, no	-, no	-, no
% of Majors in any Alliance	+, yes	+, no	+, no
% of Majors in Defense Pact	+, no	-, no	-, no
% of Major with Minor	+, yes	+, close	+, close

As for the other dependent variable, *nation-months of war begun - all*, we see the coefficient on *percent of all in any alliance* move from positive significance in the bivariate and *percent of all in a defense pact* moves from large positive almost significance in the bivariate study to large negative nonsignificance in the OLS regression. In one MLE estimation, the coefficient is positive, while in the other, it is negative. The effect of *percent of majors in any alliance,* while never significant, is positive for the bivariate correlations and negative in the other models. Again, while nonsignificant, the *percent of majors in a defense pact* is positive in the first two analyses and negative in the last two.

Table 9. ***Results of Various Estimation Procedures for the Number of Nation-Months of War Begun Regressed on Alliance Indicators (coefficient direction, conventional significance)***

	Bivariate Correlations	Ordinary Least Squares	Event Count	Generalized Event Count
% of all in any Alliance	+, yes	+, yes	-, no	-, no
% of all in Defense Pact	+, yes	-, no	+, no	-, no
% of Majors in any Alliance	+, no	-, no	-, no	-, no
% of Majors in Defense Pact	+, no	+, no	-, no	-, no
% of Major with Minor	+, close	+, no	+, close	+, yes

OLS estimations to negative nonsignificance in the event count models. The effect of The *percent of major with minor* is positive and almost significant in three out of four estimations but is not even close to significant in the OLS regression.

In the end, we have not learned much regarding the effects of alliances as measured by these variables on the onset, magnitude and severity of war. We can neither accept the Singer-Small hypothesis regarding the effect of alliance commitments on the onset of war nor can we reject the null hypothesis that alliances have no effect upon the onset of war.

Counter-Arguments

Some might argue that despite the fact that new methods and techniques might supersede older techniques, the older techniques are adequate for certain questions which would eliminate the need for constant reestimation.[25] The analogy here is to Newtonian vs. Einsteinian physics. Although Einsteinian physics works better in the extremes, Newtonian physics still works fine for high school projects involving inclines and projectiles. This argument is potentially very damaging to the analysis presented here. Linear regression is indeed robust under a wide variety of circumstances (Achen, 1982). Whether one is a frequentist or a subjectivist rarely affects the actual practice of econometric modeling (King, 1989a). The issues being studied by social scientists, however, are extraordinarily complex phenomena. As complex as our models get, the processes we study are infinitely more complex. The study of complex social events using "older" techniques is analogous to the study of time using Newtonian theory when the subject is moving at the speed of light. Old methods may work for certain questions, but not for the questions which interest us. For example, OLS estimates of models with noncategorical discrete dependent variables are plagued by large inefficiencies and nonsensical results (King, 1989b). To argue that the older methods will produce results that are "good enough" will not suffice.

Others will argue that much of the advance in statistical technique simply allows one to weaken previous assumptions, more clearly discriminate among possible alternative explanations, and improve measures.[26] That is, ignoring the measurement issue (whether one should look at large or small periods of statistical change), econometric theory and practice is basically changing monotonically in terms of truth-value. Determining the progressivity of econometric theory, however, is extremely difficult. Unlike certain

[25] Chris Achen - personal communication.
[26] John Jackson - personal communication.

sciences such as medicine, where curing sick people and increasing life expectancy is a measure of progress, the progressivity of econometrics cannot be measured by results. There is no recourse to a higher authority, no external standard.[27] Even Monte Carlo simulations are of limited use in this regard. For example, when testing linear regression with a Monte Carlo simulation, unless one chooses proxy variables with high correlated measurement error, one is unlikely to see a coefficient sign reversal.

Furthermore, Kuhn (1970) points out that, because there is no recourse to an external standard, when a non-monotonic shift in theory occurs[28], we cannot expect scientists or researchers to attest to it directly. That is, the researcher who has experienced such a shift would argue that reality had remained the same and that only the researcher's interpretation had changed. To alter Kuhn's example slightly, the scholar who has recently seen her results change due to the use of an event count model does not say, "I used to see X affect Y, but now I see X does not affect Y." This would imply that a sense in which the earlier observation had once been correct. Rather, the scholar says, "I once thought that X affected Y, but I was mistaken."[29] Researchers themselves then are of little help.

To argue that econometric theory and practice more or less moves monotonically is to argue that the rise and fall of econometric methods and practices are not historically determined.[30] Stephen Stigler (1986) argues that in the case of the Lexis ration, social-

[27] Some may argue that proofs provide econometrics with a measure of progressivity. Lakatos (1978) contends, however, that there is no consensus in the mathematical community between pure and applied mathematicians and logicians on the scientific status of proofs. He also argues that mathematical proofs are reliable although it is unclear what they are reliable about.

[28] Paradigm or gestalt shifts in Kuhn's language.

[29] The actual Kuhn example is,
"Looking at the moon, the convert to Copernicanism does not say, 'I used to see a planet, but now I see a satellite.' ...Instead, a convert to the new astronomy says, "I once took the moon to be (or saw the moon as) a planet, but I was mistaken" (Kuhn, 1970).

[30] It is simply too much of a coincidence to argue that econometric

philosophical considerations weighed heavily on the development and assessment of the test. A.W.F. Edwards (1992) argues that developments in the method of maximum likelihood were by and large ignored because the method of least squares "provided a practical procedure of such utility that logical qualms were reserved until later." Edwards (1992) further argues that the quarrels between Fisher and Pearson after 1912 were over Fisher's use of the words "inverse probability" which had been used in a different way by Todhunter. Edwards describes this episode as "one of the most influential errors of terminology in statistics." To take a more recent example, it is hard to imagine the rise of stepwise regression without the prior rise of the computer. Econometrics is therefore a matter of context as well as content (MacKenzie, 1989).

The most damaging rejoinder to this criticism, however, is that whether or not econometric practice changes monotonically is irrelevant. It still changes. Unless one is ready to posit an "end-to-statistical-history", the problem remains. The policy analyst, in particular, is not interested in the direction of change. Quite understandably, the analyst is simply confused because at time *t* alliances affect the onset of war and at time *t+1* alliances do not affect the onset of war. At time *t+2*, alliances may once again affect the onset of war. Such results are of little use to the policy planner and analyst.

Opponents of this analysis may also point to scientific fields such as medicine that appear to cope with problems similar to those of social science and econometrics. Medical analyses have changed dramatically with the advent of microscopes, x-rays, CAT scans, and MRI. Technological advance leads to changes in diagnosis and treatment. Two counter-arguments may be made here. First, controlled experiments are easier to perform in the medical sciences, which lends robustness to experimental results. Second, when it comes to health policy, medicine is often in a disarray similar to that of the social

---

theory changes monotonically *and* that the change is historically determined.

sciences. Recent, very public, disagreements over mammograms, the health benefits of vitamins, and the prevention of AIDS with condoms underscore the degree to which the medical community must rely on consensus as well as empirical evidence when making policy.

Partial Solutions

This paper seems to imply that short of continual reestimation we can have little confidence in the results of our quantitative analyses and at the most basic level, this conclusion is correct. Note, however, that the fact that changing techniques alter our results and therefore our inferences does not imply that the inferences drawn from any particular analysis are incorrect. The question then becomes how to lend robustness to results that theory leads us to believe are correct. Two methods suggest themselves. The first is to include econometric technique in specification testing. With the wealth of available statistical models, it is folly to assume that only one technique is appropriate to the question at hand. Furthermore, superseded techniques do not disappear and are still available for use. In the analyses presented here, we can say with some confidence that *percent of major with minor* has some effect on *the frequency of war begun* because the coefficient remains positive and significant (or close to it) across three different econometric models. Nothing more than this level of certainty is necessary in political science. It is usually “good enough” to say that an independent variable has some important effect on the dependent variable (King, 1986). This fix is only a partial solution because we can only include in our specifications techniques that exist. With the advent of a new technique in the future, the effect of *percent of major with minor* on *the frequency of war begun* may change.

The second method for lending robustness to our results is to combine quantitative analysis with qualitative analysis. By demonstrating how the effects posited by our statistical analyses actually work in individual cases, we may convince more readers that our results should be taken seriously. Process-tracing and structured, focused comparisons will also serve to feel out the limits of current theory and suggest revisions in our models (Achen and Snidal, 1989). Of course, qualitative analyses are also based on a theory, namely, Mill's method of agreement and method of difference. This particular theory, however, changes at a much slower rate than statistical theory. Combining quantitative and qualitative analyses will not only help convince quantitative researchers but also qualitative researchers who often dismiss out of hand the empirical generalizations generated by quantitative studies.

Conclusion

What this discussion implies about truth in political science statistical analyses remains to be addressed. Two divergent conclusions suggest themselves. On one hand, we may assume with mainstream statistics, despite the fact that econometric theory does not always change monotonically, that the event count models have a greater truth content than the OLS models and that the OLS models have a greater truth content than the correlational analysis. If this inequality is indeed the case, we must face two serious implications. One, statisticians and mathematicians will control the fate of our models and theories. Political scientists will be reduced to waiting to hear the latest word on newest appropriate statistical techniques. Two, the truth at any particular time will at best be temporary until an end-to-statistical history is reached.

The other possible conclusion argues that, due to ever-changing econometric theory and practice, we will never have confidence in the results of our quantitative analyses. One may then ask, if we cannot get at the truth through quantitative studies, why persist? This question is akin to asking an English professor why she studies literature. There is no "truth" to be found in the study of literature and it is quite common for readers to interpret works in an extremely personal manner. The English professor persists in her endeavors in order to gain insight into the human condition and insight is not the sole property of correctly specified econometric models. Even if we cannot have complete confidence in our results, that fact does not prevent us from noticing or learning something that may be of use to us in the future. Political scientists must realize that serious flaws characterize all methodologies and therefore we can ill afford to follow only one path to insight. As Teddy Roosevelt once said, “Do what you can with what you have where you are.”

Works Cited

Achen, Christopher H. and Duncan Snidal. 1989. "Rational Deterrence Theory and Comparative Case Studies." *World Politics*. 41 (January):144-169.

Achen, Christopher, H. 1985. "Proxy Variables and Incorrect Signs on Regression Coefficients." *Political Methodology*. 11 (299-316).

Achen, Christopher, H. 1982. Interpreting and Using Regression. Newbury Park, CA: Sage Publications, Inc.

Boyd, Richard. 1993. "Confirmation, Semantics, and the Interpretation of Scientific Theories," in The Philosophy of Science, edited by Richard Boyd, Philip Gasper, and J.D. Trout, pp.3-35. Cambridge, MA: MIT Press.

Campbell, Donald T. and Julian C. Stanley. 1963. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally College Publishing Company.

Chalmers, A.F. 1982. What is this thing called Science? An assessment of the nature and status of science and its methods, 2ed. Queensland: The University of Queensland Press.

Edwards, A.W.F. 1992. Likelihood: Expanded Edition. Baltimore: The Johns Hopkins University Press.

Greene, William H. 1993. Econometric Analysis, 2ed. New York: Macmillan Publishing Co.

Hanson, Norwood Russell. 1961. Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science. Cambridge: Cambridge University Press.

Hanushek, Eric A., and John E. Jackson. 1977. Statistical Methods for Social Scientists. San Diego, CA: Academic Press, Inc.

Judge, George *et al*. 1985. The Theory and Practice of Econometrics, 2ed. New York: John Wiley and Sons.

King, Gary. 1989a. Unifying Political Methodology: the Likelihood Theory of Statistical Inference. Cambridge: Cambridge University Press.

King, Gary. 1989b. "Event Count Models for International Relations: Generalizations and Applications." *International Studies Quarterly*. 33:2 (June):123-147.

King, Gary. 1986. "How Not to Lie With Statistics." *American Journal of Political Science*. 30 (August):666-687.

Kmenta, Jan. 1986. Elements of Econometrics, 2ed. New York: Macmillan Publishing Company.

Kotz, Samuel and Norman L. Johnson, eds. 1992. Breakthroughs in Statistics, Vols. I and II. New York: Springer-Verlag.

Kuhn, Thomas S. 1970. The Structure of Scientific Revolutions, 2ed. Chicago: The University of Chicago Press.

Lakatos, Imre. 1978. Mathematics, science and epistemology: Philosophical Papers, Volume 2. Cambridge: Cambridge University Press.

Leamer, E.E. 1983. "Let's Take the Con Out of Econometrics." *American Economic Review*. 73 (March):31-43.

MacKenzie, Donald. 1989. "Probability and statistics in historical perspective." *Isis*. 80:1 (March):116-124.

Singer, David J. and Melvin Small. 1968. "Alliance Aggregation and the Onset of War, 1815-1945." In Quantitative International Politics: Insights and Evidence, edited by J. David Singer, pp.247-288. New York: The Free Press.

Singer, David J. and Melvin Small. 1966. "Formal Alliances, 1815-1939: a Quantitative Description." *Journal of Peace Research*. 3:1 (January):1-32.

Stigler, Stephen M. 1986. The History of Statistics: The Measurement of Uncertainty before 1900. Cambridge, MA: Belknap Press.