Proper Nouns and Methodological Propriety: Pooling Dyads in International Relations Data

Gary King

The intellectual stakes at issue in this symposium are very high: Donald P. Green, Soo Yeon Kim, and David H. Yoon apply their proposed methodological prescriptions and conclude that a key finding in the field of international relations is wrong: democracy "has no effect on militarized disputes."[1] Green, Kim, and Yoon are mainly interested in convincing scholars about their methodological points and see themselves as having no stake in the resulting substantive conclusions. Their methodological points, however, are also high stakes claims: if correct, their claims would invalidate the vast majority of statistical analyses of military conflict ever conducted.

Green, Kim, and Yoon say they "make no attempt to break new ground statistically," but, as we will see, this both understates their methodological contribution to the field and misses some unique features of their application and data in international relations. On the latter, Green, Kim, and Yoon's critics are united: John R. Oneal and Bruce Russett conclude that Green, Kim, and Yoon's method "produces distorted results," and show even in Green, Kim, and Yoon's framework how democracy's effect can be reinstated.[2] Nathaniel Beck and Jonathan N. Katz are even more unambiguous: "Green, Kim, and Yoon's conclusion, in Table 3, that variables such as democracy have no pacific impact, is simply nonsense. . . . Green, Kim, and Yoon's [methodological] proposal . . . is never a good idea."[3]

The symposium participants deserve many thanks for putting up with my relentless questions in good humor. I also appreciate their helpful comments on an earlier draft of this article. The editors of *IO* were also very helpful. Thanks also to Ethan Katz for superb research assistance and helping me think through these issues, to the National Science Foundation (SBR-9729884, SBR-9753126, and IIS-9874747), the National Institutes of Aging (PO1 A617625-01), and the World Health Organization for research support, and to Joanne Gowa for spotting an error in an earlier draft.

1. Green, Kim, and Yoon 2001 (this issue). They also find surprising conclusions in the area of bilateral trade, but there is much less disagreement among the symposium participants about the methods involved in this issue. I shall not discuss it further.
2. Oneal and Russett 2001 (this issue).
3. Beck and Katz 2001 (this issue).

International Organization 55, 2, Spring 2001, pp. 497–507
© 2001 by The IO Foundation and the Massachusetts Institute of Technology

https://doi.org/10.1162/002081801511406667 Published online by Cambridge University Press

My given task was to sort out and clarify these conflicting claims and counterclaims. The procedure I followed was to engage in extensive discussions with the participants, including joint reanalyses provoked by our discussions and passing computer program code (mostly with Monte Carlo simulations) back and forth to ensure we were all talking about the same methods and agreed with the factual results. I learned a great deal from this process and believe that the positions of the participants are now a lot closer than it may seem from their written statements. Indeed, I believe that all the participants now agree with what I have written here, even though they would each have different emphases (and although my believing there is agreement is not the same as there actually being agreement!).

Green, Kim, and Yoon's Contribution

To understand the issues, we must separate the problem identified by Green, Kim, and Yoon from their proposed solution. The problem is unambiguous and monumentally important to this literature. It has not before been addressed in any detail, and Green, Kim, and Yoon deserve substantial credit for focusing our attention on it. I will describe the same issue in three ways:

1. Unlike, say, simple random survey sampling, dyadic observations in international conflict data have *complex dependence structures*. In a survey, observations 1 and 2 are two people who almost surely have never met and have no relationship. In contrast, in dyadic data, observation 1 may be U.S.-Iraq; observation 2, U.S.-Iran; and observation 3, Iraq-Iran. The dependence among these separate observations is complicated, central to our theories and the international system, critical for our methodological analyses, and ignored by most previous researchers. In addition, each of these dyads is observed over time, making for time-series cross-sectional data, which introduce other dependence issues.
2. An important but often unstated assumption in many statistical analyses is *exchangeability*. Roughly, this means that after taking into account the explanatory variables, one should not expect to be able to predict or explain conflict any better by knowing the names of the dyads. Since the explanatory variables normally available in international conflict data are neither powerful nor even adequate summaries of our qualitative knowledge, exchangeability is usually violated. That is, even knowing contiguity, capability ratios, growth, alliance status, democracy, trade/GDP, and lagged disputes, we would probably expect the Iran-Iraq dyad to be more belligerent than the U.S.-Mexico dyad. (Exchangeability is what enables us to use many observations to reduce our uncertainty in making a small number of inferences; it is the assumption that area studies scholars are implicitly critiquing when they point out the uniqueness of each individual case.)

https://doi.org/10.1162/002081801511406667 Published online by Cambridge University Press

3. *Unmeasured heterogeneity,* the term that describes violations of exchangeability, causes two statistical problems. At best, if this heterogeneity is unrelated to democracy (or whatever one's causal variable), then only standard errors and other assessments of uncertainty are biased.[4] This is serious, but more serious problems occur when heterogeneity is correlated with the key causal variable, and if so, estimates of the key quantities of interest will be biased. In fact, this is what is known in other contexts as "omitted variable bias." For example, suppose a degree of antipathy exists between pairs of countries, based on cultural, historical, or personal animosities, that has not been measured. (For example, completely accounting for problems between India and Pakistan by the usual list of annual dyadic variables we have measured seems unlikely.) The "historical animosity" variable is (1) unmeasured, probably (2) causally prior to and (3) correlated with democracy, and (4) affects the probability of conflict—precisely the conditions for large omitted variable biases.[5]

All the participants recognize the central importance of Green, Kim, and Yoon's methodological criticisms. Indeed, Beck and Katz write, "We close by agreeing with Green, Kim, and Yoon that the assumption of complete homogeneity of data, across both units and time, is usually suspect." Few of us regard tests such as those performed by Green, Kim, and Yoon as determinative, but they do provide some empirical evidence about the existence of heterogeneity, its correlation with democracy and the other explanatory variables, and hence the strong likelihood of bias. The issue is whether the particular approach to the problem chosen by Green, Kim, and Yoon is appropriate.

Summary of Substantive Conclusions

Green, Kim, and Yoon and, to some degree, Oneal and Russett summarize their empirical analyses with raw logistic regression results, which are difficult to interpret directly and in my view should not be attempted.[6] For the results from their pooled model, Oneal and Russett report a "first difference," which is the increase in the probability of conflict that results from a specified increase in democracy. Unlike raw logit results, the first difference is indeed of substantive interest and may even be the ultimate quantity of interest to be reported for the effects of democracy. Unfortunately, first differences, and indeed every quantity of interest but one, are impossible to compute correctly from estimates of the fixed-effects model. This is

4. Strictly speaking, this is true only if the model is linear. In versions of logistic regression, the technique of choice in this literature for binary outcome variables, coefficients can also be biased even if unmeasured heterogeneity is unrelated to the key causal variable. However, the degree of a problem will normally be considerably less with independence.

5. For example, King, Keohane, and Verba 1994, 170.

6. There is never a reason to do so; see King, Tomz, and Wittenberg 2000.

https://doi.org/10.1162/002081801511406667 Published online by Cambridge University Press

TABLE 1. *Alternative estimates of the relative risks of nondemocracy*

Analysis	*Sample frame*	*Author*	*Relative risk*
Pooled	1951–92	Green, Kim, and Yoon 2001	1.8
Pooled with dynamics	1951–92	Green, Kim, and Yoon 2001	1.7
Fixed effects	1951–92	Green, Kim, and Yoon 2001	1.0
Fixed effects with dynamics	1951–92	Green, Kim, and Yoon 2001	1.0
Pooled with dynamics	1885–1992	Oneal and Russett 2001	1.7
Fixed effects	1885–1992	Oneal and Russett 2001	1.4
Fixed-effects vector autoregression	1885–1992	Oneal and Russett 2001	1.7
Fixed effects	1951–65	King and Zeng[a]	.9
Fixed effects with dynamics	1951–65	King and Zeng[a]	.7
Fixed effects	1965–85	King and Zeng[a]	.8
Fixed effects with dynamics	1965–85	King and Zeng[a]	.8
Fixed effects	1985–92	King and Zeng[a]	6.0
Fixed effects with dynamics	1985–92	King and Zeng[a]	6.8

Note: Each entry is an estimate of the proportionate increase in the probability of military conflict resulting from a decrease in democracy from 5 to −5.

[a]Unpublished analyses by myself and Langche Zeng.

a very serious flaw in the fixed-effects model and one that I discuss further here. (The "fixed-effects model" first differences reported by Oneal and Russett in their Tables 2 and 3 are computed incorrectly, and no appropriate correction could be computed.) The one quantity of interest that can be computed from the fixed-effects model is the relative risk, which is used sometimes in international conflict studies.[7] Journalists also frequently report relative risks in medical research—for example, that the use of some drug doubles the probability of cancer. In the present context, the relative risk is the proportionate increase in the probability of conflict when the democracy variable changes from 5 (moderately democratic) to −5 (moderately undemocratic). Although relative risks were not computed by any of the participants, I have done so in Table 1 for all their quantities so that we might get some sense of the substantive conclusions resulting from their analyses.[8]

The first row of Table 1 gives the relative risk for something close to the standard specification in the literature. This figure indicates that decreasing democracy nearly doubles the probability of a dispute. More precisely, the probability of a dispute

7. Bennett and Stam 1998.

8. Since the population fraction of conflicts is very small, then $e^{(D_1-D_0)\beta_d}$ approximates the relative risk, where D_0 and D_1 are values chosen to change democracy from and to, respectively, and β_D is the corresponding logistic regression slope coefficient. I computed these relative risks from the numbers in the participants' tables and so did not compute standard errors, which requires reanalyses of their data; see King and Zeng forthcoming.

https://doi.org/10.1162/002081801511140667 Published online by Cambridge University Press

increases by 1.8 times, or, in other words, by 80 percent. The second row in the table shows that the relative risk changes relatively little when time-series dependence is taken into account. The second pair of rows in the first panel display Green, Kim, and Yoon's central substantive finding: when they add three fixed effects, the relative risk of democracy drops to 1.0, which is no effect at all. The second panel in the table portrays Oneal and Russett's results. They first extend the time series back to 1885, run the classic pooled analysis, and find an effect approximately like the same effect in Green, Kim, and Yoon's shorter time series. In the longer time series, their fixed-effects regression causes the 1.7 relative risk to decline only to 1.4 rather than to be eliminated entirely. And with dynamics through their vector autoregression approach, relative risk recovers its original value. Dropping such a large fraction of the observations increases the inefficiency (variance) of the fixed-effects results; Oneal and Russett's approach recovers some of this lost variance with additional observations. The cost here, of course, is a much stronger, and more difficult to defend, exchangeability assumption. Of course, Oneal and Russett do not trust the fixed-effects analysis, with or without the longer time series, and only performed these analyses to show that the effect for democracy could even be recovered in that more difficult, and perhaps inappropriate, context.

Finally, I make Green, Kim, and Yoon's point in another way by examining their model in different subperiods of their data.[9] The last panel of Table 1 shows that in the periods 1951–65 and 1965–85 there is essentially no effect of democracy (the point estimate indicates that decreasing democracy even slightly reduces the probability of conflict), but the relative risk is much larger in the period 1985–92: the probability in that period increases by a factor of six or more when democracy decreases.

Taken together, these results indicate very substantial unmeasured time-series heterogeneity. Some additional analyses conducted by Oneal and Russett on their data, resulting from our discussions of these results, provide some additional support for an increasing effect of democracy over a longer period of time. However, their research and mine (not shown) indicate that using the pooled model, the effect of democracy seems much more, even though not entirely, stable. This is an important topic for future research: if the international system has changed in this massive a way, can we identify the substantive variable that accounts for the change so that exchangeability still holds and we can still use all the data to draw our inferences?

The results in Table 1 are of some interest, but relative risks in all fields in which they are used are regarded as inadequate for understanding statistical results. For example, a relative risk of 2 could summarize a change in probability from 1 in a billion to 2 in a billion or from 0.4 to 0.8. In other words, the same relative risk could indicate a result that is substantively irrelevant or vitally important. The only way to know the difference would be to estimate a first difference, marginal effect, or the

9. Langche Zeng and I found these results when looking for an example for a different paper. We used the replication data set made available by Green, Kim, and Yoon, which ensures that we had the same data as the participants.

https://doi.org/10.1162/002081801511406667 Published online by Cambridge University Press

base probabilities of conflict given specific configurations of the values of the explanatory variables. Unfortunately, with the methods offered by Green, Kim, and Yoon, these more interesting quantities cannot be computed.

Methodological Evaluation

If we had a good measure of the omitted variable that caused the unmeasured heterogeneity, such as historical animosity, and it has the characteristics I describe in the first section, including it in the usual pooled logit equations as an extra variable would greatly improve our analyses. In addition, measuring this variable, or whatever is the substantive variable underlying the unmeasured heterogeneity, is by far the best strategy to address the problem at hand.

Unfortunately, apart from new data collection efforts, our options are normally quite limited when it comes to omitted variable bias. Remarkably, however, information about the unmeasured heterogeneity can often be gleaned from time-series cross-sectional data, and corrections can sometimes be made, without new measurements. That is the promise of Green, Kim, and Yoon's fixed-effects regressions: The theory is that by controlling for a set of dyad-level indicator variables, all dyad-specific heterogeneity is controlled, including otherwise unmeasured variables such as historical animosity. The intended result is that with the omitted variable effectively in the analysis, the bias would vanish.

Suppose the only potential problem for an analysis is unmeasured heterogeneity. Whether the outcome variable is continuous or binary, a very large number of informative observations for each cross-sectional unit (not merely a large number of observations) is sufficient to ensure that including fixed effects will be an improvement over the usual pooled binary logit model.[10] In neither the continuous nor the binary case will fixed effects necessarily be an optimal approach, even though it will normally help remove some of the omitted variable problem as compared to ordinary pooled logit.

The issue is whether the fixed-effects model works in the present case, which does not meet these criteria. In (binary) rare events data like these, the amount of information in the data depends not only on the number of observations but also on the rareness of events.[11] Although Green, Kim, and Yoon have (at most) forty-two observations per dyad, each contains very little information. Indeed, the dependent variable is a constant for most of the dataset: 2,877 of the 3,075 dyads have no disputes at all and so have all-zeros for every annual observation. Of the 198 remaining dyads, 116 have only one dispute (a string of zeros with only a single one).

10. Beck and Tucker 1996.
11. King and Zeng forthcoming.

https://doi.org/10.1162/002081801511140667 Published online by Cambridge University Press

Indeed, as it turns out, the all-zero dyads, and other rare events problems, are at the center of the present controversy. The issue is that the fixed-effects model is "inestimable" with all-zero dyads. A model that is inestimable (also known as "not identified") cannot be estimated no matter how good the data are. The problem is that the dyad-level indicator variables corresponding to the all-zero dyads perfectly predict the zeros in the outcome variable. And although it might seem that perfect prediction is one of those problems that political scientists would love to deal with, it wreaks havoc with the logit model.

Consequently, one needs to take some other action, and there are two possibilities, depending on how you conceptualize the data and model. The differences between these strategies (each of which is a different estimator of the same model) are small in practice, but explicating the differences helps in understanding the problems with both.

Perhaps the more intuitive method is known as the *fixed-effects logit estimator*. The idea here is to drop the all-zero dyads and corresponding indicator variables and run the analysis on the observations and variables that remain. When there is a lot of information in each dyad that remains, the coefficients on democracy and the other substantive variables are estimated consistently, although the coefficients on the remaining indicator variables are biased. This strategy is not satisfactory for several reasons. Not only is there bias in estimates of the remaining indicator variables and, of course, no estimates on the excluded indicator variables; there is also probably not enough information in what remains to estimate the coefficients on democracy (and the other substantive variables) well. Unless the number of time periods were much larger and/or events were much less rare, this estimator would produce statistically inconsistent estimates of all slope coefficients. In addition, consistent estimates of all the coefficients on all the variables (including those omitted) are required in order to compute quantities of interest other than the relative risk, and so the fixed-effects logit model in the presence of all-zero dyads cannot get us what we need.

The other strategy, which was adopted by Green, Kim, and Yoon and Oneal and Russett, is *Chamberlain's logit estimator*, also sometimes known as *clogit*.[12] This procedure works by giving up entirely the goal of estimating the coefficients on the indicator variables. (Clogit estimates the logit coefficients in the same model as the fixed-effects logit model, but it is a different estimator, requiring a specialized computational procedure, such as exists in Stata.) Clogit conceptualizes the problem by asking in each dyad whether there will be a dispute in each year, assuming knowledge of how many disputes there were during the entire observation period. Assuming knowledge of the future (that is, "the entire observation period") to understand the past in this way is hard to justify, even if the goal has nothing to do with forecasting. Although clogit makes sense in other applications (such as two spouses in each of many families), the present application really does not fit the

12. Chamberlain 1980.

https://doi.org/10.1162/002081801511406667 Published online by Cambridge University Press

theory well. And, of course, the lack of estimates on the indicator variables means that no quantity of interest other than relative risk can be computed. The all-zero dyads get dropped in this approach as well, since once one knows that no conflicts have occurred, there is no uncertainty about whether a conflict occurred in any one year.

In both approaches, dropping the all-zero dyads is thus an expedient approach from a methodological perspective enabling one to estimate at least some of the necessary parameters, but the procedure can be interpreted from a substantive perspective as well. Oneal and Russett explain the dilemma: "It is simply impossible to think that the 97,150 annual observations of the experiences of the 2,751 dyads that managed to live in peace—84 percent of our total number of cases—tell us nothing about the causes of war." Unfortunately, this assumption is a consequence of choosing either the clogit estimator or the fixed-effects logit estimator. If you regard the dyadic indicator variables as a causal consequence of democracy (and the other substantive variables), then the model assumes that there are no substantive explanatory variables that could ever account for why the U.S.-Canada dyad is at peace or why it is more at peace than the Iran-Iraq dyad. If, however, you regard the dyadic indicators as causally prior to democracy, then the model assumes that it is impossible to identify a substantive variable that intermediates between, and thus accounts for, the indicator variables and peace. Either way, the famous comparative politics dictum of "getting rid of proper nouns" is not only something that was not achieved in the Green, Kim, and Yoon analysis, but is also impossible to achieve under the proposed model, no matter how much our data collection efforts improve. In their concluding section, Green, Kim, and Yoon appropriately suggest avoiding this problem altogether by searching for better substantive covariates that would enable researchers to control for the heterogeneity without the dyad-level indicator variables. Getting better data is usually the best advice, and it clearly is here. Green, Kim, and Yoon also suggest a sequence of tests and procedures that might aid in this goal.

Concluding Suggestions: Clean Pools of Salamanders

So where are we? Green, Kim, and Yoon have identified unmeasured heterogeneity as a critical methodological problem that has not been addressed previously. Their proposed solution is not really adequate to the task at hand, even though it served them well in demonstrating how much the standard results change when they alter some implausible assumptions the field has taken for granted. So we have a problem and no solution. That is a problem for international relations, but an important opportunity for some enterprising methodologists out there. I conclude with some suggestions for researchers in international relations doing research now, and for methodologists working to improve future applied conflict research.

https://doi.org/10.1162/002081801511406667 Published online by Cambridge University Press

Suggestions for Conflict Researchers

Since they can get better data, substantively oriented researchers almost always have better tools to solve methodological problems than methodologists. They only need remember that sufficiently good data beats better methods every time. This is an important aphorism for conflict research and one that researchers in this field seem to understand, at least to a degree. Indeed, over the last several decades, considerable effort has gone into cataloging and categorizing every manner of international dispute.

Unfortunately, even though our databases have hundreds of thousands of observations, they still contain relatively little information, making inference difficult. The low information content of our data does not necessarily indicate that our efforts are flawed, since data collection strategies cannot create information where none exists. The rareness of international disputes is merely a fact about the world that we need to cope with. Surely investing additional effort into refinements in definitions and measurements of "militarized interstate disputes" and other similar concepts seems wise, but there is a limit to this strategy. More fertile ground for learning about international conflict is better found in trying to improve our set of measured explanatory variables. The covariates available to explain and predict international conflict, both those that are the subject of causal inference and those used as control variables, are very crude measures of underlying constructs, and they exclude many concepts altogether. The split between quantitative and qualitative researchers may be more severe in this field than in any other in political science, and the validity and comprehensiveness of our explanatory variables may be the most important reason.

Green, Kim, and Yoon's argument about unmeasured heterogeneity (or equivalently, about omitted variable bias) can be solved easily with better measures of more appropriate control variables. Indeed, redoubling our efforts to find these measures would be the most important action conflict researchers could take in response to this symposium. If the new measures are not available (yet), then we must recognize that pooled analyses risk omitted variable bias. Unfortunately, fixed-effects models in the context of rare events data, like those in international conflict, do not enable us to apply a methodological fix to get around this omitted variable bias problem. But this does not let anyone off the hook, since bias does not vanish just because we lack a solution.

Fortunately, even when we are convinced of the potential for omitted variable bias, and have some idea what the omitted variable is, but we have no measure of it, some action can still be taken. As shown by King, Robert O. Keohane, and Sidney Verba, the direction of the bias can be ascertained.[13] To continue with my running example, suppose democracy is the key explanatory variable and historical animosity is the omitted explanatory variable. In the linear case (which is usually appropriate to use as an analogy for the logit case, even though it is not exact and can be wrong in some instances), instead of estimating the effect of democracy, by

13. King, Keohane, and Verba 1994.

https://doi.org/10.1162/002081801511406667 Published online by Cambridge University Press

omitting historical animosity we are actually estimating that effect plus a bias term. This bias term is the product of two factors—the correlation between democracy and historical animosity, and the effect of historical animosity on the probability of a dispute. Even though we do not have a measure of this omitted variable, it seems reasonable to conclude that dyads with high levels of historical animosity have lower levels of democracy, and so the correlation is negative. Similarly, it is likely that increasing historical animosity produces a higher probability of a conflict. Thus, the bias is the product of a negative number and a positive number and so is itself negative. This means that instead of estimating the effect of democracy, which is hypothesized to be negative, we are actually estimating the effect plus a negative bias term. This, in turn, means that we are estimating something too small (that is, too large a negative number) or, in other words, that democracy reduces the effect of war less than indicated by the pooled analysis. Put in yet another way, this means that if we had measured and controlled for historical animosity, and our assumptions are correct, then the effect of democracy would be smaller than indicated by the pooled analysis.

Of course, this example only gives a taste of the kinds of analysis that could be done even without better measures. A full analysis in the context of a real application should follow the main point learned from this symposium and systematically search out and document what the omitted variables are. At best, they should then be measured and controlled for. At worst the direction of the bias induced should be ascertained and reported. Until better data or improved methods are available, it is hard to see why we should not expect this type of work to accompany every subsequent analysis using international conflict data.

Suggestions for Political Methodologists

A logical methodological starting point for addressing the problems at hand would be based on Bayesian hierarchical, random effects, or split population models. Beck and Katz cite several examples of these models. Just like the fixed-effects logit model, these represent compromises between the extreme of the pooled logit model and the "equation-by-equation" extreme where a separate analysis would be run on the time series in each dyad. Unlike fixed effects, all these analyses are probabilistic. They borrow strength statistically from similar dyads to help estimate the quantities of interest in each one. An advantage of these approaches is that they should not require dropping the all-zero dyads.

However, the standard hierarchical models in this area have two features that should be changed. First, most models assume that the unobserved, but estimated, heterogeneity is independent of democracy (and the other substantive variables). Assuming independence would assume away the omitted variable bias problem and would fix nothing. Fortunately, it is not difficult to change this assumption, but it must be done.

Second, an approach that extracts the most information will likely be one that directly models the unique structure of dyadic data. Unfortunately, no off-the-shelf

https://doi.org/10.1162/002081801511406667 Published online by Cambridge University Press

model is available for these data, but there is a close analogy in the statistical literature on salamander mating experiments that might help a methodologist build one. These researchers isolate each pair of male and female salamanders and see whether they mate, and they repeat the process for all possible pairs. The structure of the data is quite similar: dyadic data with complex dependency structures and with a binary outcome variable. Although there are some methodological differences (countries are not male and female, events in salamander mating are not as rare as in international conflict, and the longer time series in conflict data must be considered), but the structure of the necessary statistical models are quite similar. In fact, some of their models have a structure similar to that suggested by Nathaniel Beck and Richard Tucker.[14]

References

Beck, Nathaniel, and Jonathan N. Katz. 2001. Throwing Out the Baby With the Bath Water: A Comment on Green, Kim, and Yoon. *International Organization* 55 (2):487–95.

Beck, Nathaniel, and Richard Tucker. 1996. Conflict in Space and Time: Time-Series Cross-Sectional Analysis with a Binary Dependent Variable. Paper presented at the 92nd Annual Meeting of the American Political Science Association, San Francisco.

Bennett, D. Scott, and Allan C. Stam, III. 1998. Theories of Conflict Initiation and Escalation: Comparative Testing, 1816–1980. Paper presented at the 39th Annual Meeting of the International Studies Association, Minneapolis.

Chamberlain, Gary. 1980. Analysis of Covariance with Qualitative Data. *Review of Economic Studies* 47 (1):225–38.

Chan, Jennifer S. K., and Anthony Y. C. Kuk. 1997. Maximum Likelihood Estimation for Probit-Linear Mixed Models with Correlated Random Effects. *Biometrics* 53 (1):86–97.

Drum, Melinda L., and Peter McCullagh. 1993. REML Estimation with Exact Covariance in the Logistic Mixed Model. *Biometrics* 49 (3):677–89.

Green, Donald P., Soo Yeon Kim, and David H. Yoon. 2001. Dirty Pool. *International Organization* 55 (2):441–68.

Karim, M. Rezaul, and Scott L. Zeger. 1992. Generalized Linear Models with Random Effects: Salamander Mating Revisited. *Biometrics* 48 (2):631–44.

King, Gary, and Langche Zeng. Forthcoming. Logistic Regression in Rare Events Data. *International Organization* 55 (3).

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, N.J.: Princeton University Press.

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. Making the Most of Statistical Analyses: Improving Interpretation and Presentation. *American Journal of Political Science* 44 (2):347–61.

McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*. 2d ed. London: Chapman and Hall.

Oneal, John R., and Bruce Russett. 2001. Clear and Clean: The Fixed Effects of the Liberal Peace. *International Organization* 55 (2):469–85.

Signorino, Curtis. 1999. Strategic Interaction and the Statistical Analysis of International Conflict. *American Political Science Review* 93 (2):279–97.

14. Beck and Tucker 1996. See, for example, McCullagh and Nelder 1989; Chan and Kuk 1997; Drum and McCullagh 1993; and Karim and Zeger 1992. For a different approach that might also be productive, see Signorino 1999.

https://doi.org/10.1162/002081801511140667 Published online by Cambridge University Press