International Studies Quarterly (2018) **62**, 892–902

Temporal Dependence and the Sensitivity of Quantities of Interest: A Solution for a Common Problem

RESEARCH NOTE

LARON K. WILLIAMS
University of Missouri

Scholars of international relations increasingly use temporal dependence variables (polynomials or splines) to control for unmodeled duration dependence in nonlinear models (such as logit or probit) of events ranging from interstate conflict and civil war to sanctions imposition and trade agreements. I identify two inferential obstacles that are widespread to nonlinear models, and are exacerbated by the unique features of temporal dependence variables. First, compression causes the quantities of interest to be sensitive to the values in the counterfactual scenario (most notably, time). Second, presenting substantive effects calculated at one simulation scenario (such as an "average" scenario) grossly inflates the representativeness of that scenario and neglects the variability within the sample. The consequences of these problems range in severity from understating the magnitude of the substantive effects to deriving inferences that are wholly unrepresentative of the data. I offer a simple checklist. First, use the values observed in the data to generate in-sample quantities of interest. Second, plot those quantities of interest across the offending variable (for example, time) and interpret the relationship. Finally, provide a sense of the sample's variability in quantities of interest through simple summary statistics (such as mean, standard deviation, and range). These simple fixes provide much-needed transparency and act as a shield against scholars who might otherwise present misleading results.

Introduction

Temporal dependence variables (TDVs) are nearly universal in binary models of international relations. Even though empirical tests are underpinned by theories often derived from decades of careful formal theorizing, there usually remains unmeasured and unobservable characteristics that change the underlying risk of an event occurring. In these cases, binary models are plagued by duration dependence, which is "the extent to which the conditional hazard of the event of interest occurring is increasing or decreasing over time" (Zorn 2000). Scholars wanting to avoid the problems associated with serially correlated errors have looked to duration modeling techniques to correctly model this duration dependence. Beck, Katz, and Tucker (1998) occupy a central place in this movement, and their intuition that "[binary time-series cross-section] data are grouped duration data" triggered a breakthrough in how scholars deal with temporal dependence.

As a result, TDVs—dummy variables, time counters, cubic polynomials, cubic splines, etc.—are widespread (for a review, see Carter and Signorino 2010). Many use these techniques to address a wide range of topics in international relations; in examining the imposition of sanctions (Hafner-Burton and Montgomery 2008), causes of genocide (Harff 2003), the recurrence of civil war (Walter 2004), trade liberalization (Milner and Kubota 2005); in studies of the effects of climate on communal conflict (Fjelde and von Uexkull 2012); compliance with international law (Simmons 2000) and treaties (Von Stein 2005), among others.

While helpful in controlling for unmodeled duration dependence, TDVs potentially derail inferences from nonlinear models (such as logit or probit) by complicating the interpretation of quantities of interest (QIs)—any transformations of the coefficients, such as probabilities, partial effects, risk ratios, etc.—in two ways. First, TDVs are one of the most egregious sources of inferential problems related to compression. Compression means that QIs in nonlinear models depend not only on the coefficients from the model, but also on the location of the observation along the cumulative density function (CDF). As a consequence, QIs in nonlinear models may be small or large depending on the values of other variables in the simulation scenario and not just the variable of interest. Scholars generally view TDVs as atheoretical, so they make little effort to see how time shapes the outcome. TDVs typically have large coefficients (by soaking up unmodeled temporal effects) and wide ranges of values. As a result, QIs–and the basis of the inference–are highly sensitive to the values selected for the scenarios, which causes potentially misleading inferences from substantive effects while understating the actual variability of those effects in the sample.

Second, the dominant approach to calculating QIs with the "average-case" approach (based on means, medians, and modes; see Hanmer and Kalkan 2013) greatly inflates the risk of misleading inferences. Scholars often set up "average" scenarios that unknowingly contain unrepresentative values or values that are inconsistent across the TDVs. In addition to heightening the risk of extrapolation, having multiple defensible average scenarios—without

Laron K. Williams is an associate professor of political science at the University of Missouri. His research focuses on political methodology, foreign policy public opinion, and voting behavior.

Authors' note: Previous versions of this article were presented at the "Robustness Testing and the Empirical Analysis of Observational Data," Cologne Center for Comparative Politics, the 2016 Annual Meeting of the European Political Science Association and the University of Texas Department of Government. The author would like to thank Patrick Brandt, David J. Brule, Scott Cook, Justin Esarey, Eric Neumayer, Mark Nieman, Thomas Plumper, Stephen Quackenbush, Toby Rider, Zeynep Somer-Topcu, Vera Troeger, Guy D. Whitten, Chris Wlezien, and two anonymous reviewers for their invaluable comments. Clinton Swift provided excellent research assistance.

Williams, Laron K. (2018) Temporal Dependence and the Sensitivity of Quantities of Interest: A Solution for a Common Problem. *International Studies Quarterly*, doi: 10.1093/isq/sqy036

full transparency—increases the potential for cherry-picking scenarios.

In a brief survey of these efforts in practice, I uncover a substantial gap between how methodologists understand QIs and how political scientists in general depict them. The survey shows widespread use of TDVs in studies of international relations, but a lack of interest in interpreting the effects of time. In the vast majority of studies, scholars do not enable readers to determine how sensitive the QIs are due to time (over 83.0 percent only calculate QIs based on one scenario). Moreover, an in-depth meta-analysis of twenty studies reveals additional problems that arise when using the average-case approach in the context of temporal dependence. Scenarios often contain extreme or inconsistent values of time, so scholars may report QIs that are substantially under- or overstated relative to the in-sample QIs. The meta-analysis also reveals that TDVs—when statistically significant—are consistently among the most serious offenders. These patterns suggest a variety of drawbacks related to the average-case approach and question its value in generating accurate inferences.

I develop a checklist for the interpretation and presentation of QIs that overcomes these inferential obstacles. In models where one fears that the QIs are particularly sensitive (due to TDVs, fixed effects, or spatial variables), scholars should present the distribution of in-sample QIs across the offending variable(s). Given that scholars are increasingly interested in time as a confounding variable, this ensures that scholars are aware of how QIs wax and wane across time. In addition to being based only on configurations of the independent variables that are actually observed in the sample (thus avoiding the dangers of extrapolation), providing additional information such as the standard deviation and range gives an idea of the dispersion of the QIs in the sample. This strategy emphasizes transparency and, as a result, produces inferences that are more accurate and representative of the bulk of observations. Toward this end, I provide simple Stata code in the supplementary files that automates the calculation of these values.

I then illustrate these problems in practice by exploring three recent examples that use TDVs to address unmodeled duration dependence on nuclear proliferation (Way and Weeks 2014), civil war onset (Cunningham 2013), and wartime fiscal policy (Flores-Macias and Kreps 2013). In all three cases, examining the variability of substantive effects across the TDVs adds considerable depth and nuance to their inferences. The first illustration (Way and Weeks 2014) shows that substantive effects of key variables are extremely sensitive to the values of the TDVs in the simulation scenarios, which suggests an important moderating effect of time in the relationship. A closer examination of Cunningham (2013) reveals additional obstacles. Scholars have wide latitude in selecting average values of the TDVs (for example, mode, median, and mean) in the simulation scenarios, which means that the partial effects presented by scholars can vary widely from miniscule to quite large. The third illustration reveals that the use of the average-case approach when employing complex fixes for temporal dependence (such as splines) exaggerates the risk of extrapolation and inaccurate inferences (King and Zeng 2006). I show that these inferential obstacles are present even in the case where the TDVs fail to meet conventional standards of statistical significance.

Interpretive Pitfalls in Nonlinear Models

A recent trend is to use software (such as Clarify in Stata and Zelig in R) to produce meaningful QIs that ease readers' interpretive burdens. Others extend this trend to dynamic models (de Boef and Keele 2008; Williams and Whitten 2012; Philips, Rutherford, and Whitten 2016; Williams 2016). Though certainly important in linear models (such as ordinary least squares (OLS)), providing meaningful QIs in nonlinear models is necessary to fully understand a result's substantive significance.

Econometricians have long noted that QIs in binary models are sensitive to the values of the other variables (for example, Maddala 1983; Nagler 1994; Gujarati 2003). The coefficient (β_X) merely represents the marginal effect of X on $X\beta$. The principle of compression (Berry, DeMeritt, and Esarey 2010; Rainey 2016) means that the marginal effect of $X\beta$ on $\Pr(Y)$—which is often the substantive effect of most interest—depends on the location along the cumulative density function (CDF). The result is that the marginal effect of X on $\Pr(Y)$ is nonconstant, and the marginal effect of the variable decreases as the probability shifts away from 0.5 in either direction. Compression also produces considerable sensitivity in QIs across the sample; an identical shift in X potentially produces N different QIs, depending on the observation's location along the CDF.

Because of the nonlinear nature of the coefficients, some caution is required in fully interpreting these models. Hanmer and Kalkan (2013) review two methods for calculating QIs in nonlinear models. The first method is the average-case approach, in which the scholar calculates QIs for a scenario where the other independent variables are held at their means or modes (or other average values). The second method is the *observed-value* approach, which calculates QIs given the observed values for each observation in the sample and then averages those effects across the entire sample (or meaningful subgroups) (Hanmer and Kalkan 2013).

The choice between these two methods is not without consequence, as the average-case approach has some clear drawbacks. First, the average values of the independent variables do not provide an unbiased estimate of the average probability (Train 2009). Second, Hanmer and Kalkan (2013) stress that scholars typically lack fine-grained theories about particular cases, so the effects for one scenario cannot appropriately test one's theory. Third, there is also a risk of selecting a counterfactual scenario that is far from the bulk of the sample observations (King and Zeng 2006).

Interpretation in the Context of Temporal Dependence Variables

Those using QIs to interpret nonlinear models in the context of temporal dependence must be aware of two problems that potentially derail inferences. First, TDVs are one of the most egregious sources of compression effects in political science.[1] As the survey below will illustrate, TDVs are nearly ubiquitous in quantitative studies of international relations. Since they are typically used as a methodological fix for the problem of temporal dependence, they are treated as an afterthought, and scholars often omit the coefficients. The use of splines (and polynomials to a lesser extent) mean that $X\beta$ is a nonlinear function of t, which makes it difficult to interpret based on the coefficients. Moreover, it is often the case that TDVs have substantively large coefficients as a result of soaking up unexplained temporal variance, and they have ranges that can far exceed those of other variables—Carter and Signorino (2010) call this "numerical instability." As a result, TDVs potentially determine the sensitivity of QIs

[1] Another complication is that the presence of TDVs raises the potential for a probabilistic long-term effect (Williams 2016).

by influencing the location along the CDF (I return to this point in the meta-analysis below).

For example, consider the often studied outcome of international disputes. The patterns of duration dependence in international conflict are typically negative, indicating that the baseline probability of a dispute is highest immediately following a dispute and then declines with the passage of time. With this in mind, it is clear that the QIs are quite sensitive to the scenario chosen (in this case, t) and vary across the observations in the sample. One can deduce that the largest QI will occur at extremely small values of the TDVs (since this configuration of values pushes the probability toward 0.5) and decrease at higher values of TDVs (as the probability approaches 0).

Second, using the average-case approach in the context of temporal dependence greatly inflates the risk of misleading inferences. The average-case approach may give the appearance of being an average effect (because of the use of mean, median, and modal values), but it may not in fact be representative of the sample. Furthermore, it is not clear which statistic (mean, median, or mode) is most appropriate to reflect the average value of the TDV.[2] Scholars might therefore be tempted to choose QIs from various scenarios to depict the largest substantive impact. A lack of transparency by scholars in practice means that the threat of this type of deception is very real. TDVs are often poorly understood (see Carter and Signorino 2010), which increases the risk of establishing a counterfactual scenario where the time values are meaningless, or the values are not consistent across the TDVs. The average-case approach raises the possibility of establishing a scenario that does not actually appear in the sample (King and Zeng 2006), or *extrapolation*, which increases the risk that the inferences one makes will be susceptible to changes in model specification (King and Zeng 2006; Hanmer and Kalkan 2013). Even selecting counterfactuals that are reasonable representations of the data (that is, not extrapolated) does not inoculate the QIs from being sensitive to the scenario chosen, and it certainly does not reveal the variability across the sample.

In the next section I briefly survey a sample of published work in international relations using TDVs to see if applied political science reflects our understanding of the sensitivity of QIs.

Survey

I conduct a survey of all the articles published in six political science journals[3] that cited Beck et al. (1998) or Carter and Signorino's (2010) foundational pieces on temporal dependence. For each article, I note the rationale used to justify the specification, the presentation style of the TDVs, the manner of interpretation (if any), and how the QIs were interpreted. The first point to note is the widespread acceptance and implementation of these suggestions. Out of 401 substantive articles that cite either article, 307 (76.6 percent) employ some version of TDVs in their primary empirical model.

Table 1. Summary of patterns in survey

	Yes	*No*
Hazard Rate		
Interpretation	36.2%	63.8%
Omit coefficients	59.7%	40.3%
Methodological justification	95.4%	4.6%
Theoretical justification	15.6%	84.4%
Quantities of interest	77.7%	22.3%
Simulation scenario		
Scenario information	73.4%	26.6%
Scenario information (TDV)	69.6%	30.4%
Multiple scenarios	16.9%	83.1%

Note: Cells are based on the 307 articles (out of 401) that cited either Beck et al. (1998) or Carter and Signorino (2010) that employed some version of TDVs.

The survey reveals two notable patterns in interpretive methods in nonlinear models (summarized in Table 1). First, scholars express little interest in interpreting the effects of time in their models. A little over a third of the studies interpret the hazard rate, either in text, in a table, or graphically. Moreover, nearly 60.0 percent omit at least some of the TDV coefficients entirely from their tables of estimates, which makes it impossible to infer the shape of the hazard, let alone its effects on the QIs from compression.[4] Scholars very rarely make a concerted effort to interpret the effects of these TDVs on the outcome itself and, indeed, often make choices that block the reader's ability to make those inferences on her own.

Second, current practice prevents readers from inferring the variability of QIs in the context of temporal dependence. Given that the vast majority of scholars justify the inclusion of TDVs based on methodological concerns (95.4 percent) rather than theory (15.6 percent), it is understandable that the focus is on the substantive effects of the other independent variables. The majority (77.7 percent) use QIs to provide a deeper substantive understanding of the models beyond hard-to-interpret coefficients, and almost three-quarters (73.4 percent) of those provide a broad description of the scenarios used to generate those QIs. Yet the inclusion of TDVs in these models complicates this process, so it is informative to examine how scholars treat the TDVs in these simulation scenarios. Without that information, the reader has no idea how rare or common the scenario is or whether the baseline probability of the event is closer to 0, 0.5, or 1. Of those who generate QIs and describe the simulation scenarios, 77.0 percent select some measure of central tendency (such as the mean, median, or mode) of time, while 10.3 percent select a specific, nonaverage value. Common practice is to ignore the compression effects of TDVs on QIs; in the vast majority of studies (83.1 percent), scholars calculate QIs based on only one scenario. By doing so, scholars ignore the sensitivity in effect size across the sample that arises due to compression (potentially due to the TDVs).

This survey reveals that empirical work published in top general and international relations journals has not caught up to the advice by political methodologists (Hanmer and Kalkan 2013), which has resulted in behaviors that inhibit the complete interpretation of QIs in nonlinear models.

[2] Setting the variables to their means will be especially problematic if there are extreme values of t because the $X\beta$ might trend toward extremely negative or positive values. This results in QIs that are much smaller than the other scenarios. Since natural cubic splines are intended to smooth this relationship (even at extreme values), they are less susceptible to this problem than polynomials (Keele 2008).

[3] The six journals include the *American Journal of Political Science* (through January 2017), the *American Political Science Review* (through February 2017), *International Organization* (through Summer 2017), *International Studies Quarterly* (through March 2017), *Journal of Conflict Resolution* (through November 2017), and *Journal of Politics* (through July 2017).

[4] The tendency to omit the coefficients is stronger for splines (67.9 percent) than cubic polynomials (45.4 percent), presumably because of the difficulty in interpreting the coefficients.

Meta-Analysis

While it is clear from the survey in the previous section that scholars do not fully appreciate the issues that arise when interpreting QIs in these models, it is not yet clear whether this neglect changes the inferences. To this end, I randomly select twenty articles from the above survey that generated QIs with the average-values approach based on nonlinear models with TDVs. I then calculate QIs based on the checklist described below. This meta-analysis illuminates common practices regarding the use of TDVs and reveals that the interpretive pitfalls are in fact quite meaningful.[5]

The first task of the meta-analysis is to summarize how scholars treat TDVs in their simulation scenarios. Table 2 identifies the value of the counter variable, or t, (for example, the *peace years* variable in conflict studies), its percentile in the sample (in parentheses), and whether the values are consistent across the TDVs. For example, Fuhrmann and Sechser (2014) hold their counter variable at an extremely low value (zero years, which is the second percentile), yet hold their other TDVs (cubic polynomials) at the appropriate values. On the other hand, Salehyan (2008) selects a scenario for which the counter is near the center of the distribution (twenty-eight years, which is the sixty-second percentile), but holds the three spline variables to values that are associated with different years of the counter variable (thirty-nine, forty-one, and forty-four years). Indeed, half of the studies set their TDVs to values that are not consistent or realistic. Both problems—setting t to extreme and/or unrealistic values—threaten the validity of the inferences.[6] The difficulty in establishing scenarios where the TDVs are typical and consistent illustrates a heightened risk of extrapolation associated with the average-case approach; only 10.0 percent of the baseline scenarios were in the convex hull, and a large percentage had less than 10.0 percent of the sample observations near the scenario.[7] The result is that scenarios based on the average-case approach are rarely typical of those found in the real world.

The second task of the meta-analysis is to characterize whether the QI is representative of the QIs in the sample. In Table 2 I provide the QI—changes in predicted probability[8]—from each study, its percentile (in parentheses), and its range of values. For example, generating QIs based on an extreme scenario has the potential to drastically understate (for example, Weeks 2012) or overstate (Wu 2015) the effects relative to the in-sample effects. It is clear that some scholars depict QIs that are at the extremes of in-sample QIs, a problem that is exacerbated by only using one scenario that might not be representative of the sample. It should also be noted that since these QIs are based on the average-case approach, they cannot adequately portray how much the QIs vary in the sample.

When these problems are considered in tandem, the question then becomes the following: to what extent are TDVs driving this increased sensitivity? To answer this question, I calculate the mean QI at each unique value of the IVs (for the continuous IVs I establish twenty equally sized bins) and then calculate the absolute value of the differences between the largest and smallest average. Based on these values, I rank the IVs to determine which ones cause the most sensitivity in the QIs. As shown in the final column of Table 2, the TDVs are unique in that they are consistently among the most impactful independent variables. Those few times in which the TDVs are not in the top three (such as Miller 2014; Bapat et al. 2016; Bapat and Zeigler 2016) are also those where none of the TDV coefficients are statistically significant at conventional levels (indicative of a lack of temporal dependence). This meta-analysis makes clear that, if the TDVs are statistically significant, then they are likely to generate increased sensitivity of the QIs. Even in those scenarios when the TDV is set to a reasonable value near the bulk of the distribution, the outsized influence of the TDVs heightens the potential for misleading inferences.

Recommendations

In this section I offer a series of recommendations for scholars seeking to avoid these interpretive pitfalls.[9] Models with TDVs result in QIs that are sensitive to the scenario and often have a great deal of variation across the sample. Additionally, the complex and often misunderstood nature of splines increases the chance of creating a scenario that requires extrapolation. For the reasons outlined above, the observed-value approach offers considerable improvements over the average-case approach.

I present the following checklist for generating QIs in nonlinear models with TDVs:

1. Begin by calculating the QI of X on Y given each observation's profile of covariates ($\boldsymbol{X}_i$): $\gamma_i = \Delta\Pr(Y = 1|\boldsymbol{X}_i,\ \Delta X)$. Any reasonable change in the independent variable, such as 1-unit or 1-standard deviation, is appropriate.
2. Plot the QI against the offending variable (especially t); if there is clear visual evidence of sensitivity (or a relationship), present the plot in the manuscript and interpret. If there is no clear evidence of a relationship, this information can be relegated to a footnote or appendix.
3. Present the variability of the in-sample QIs with simple summary statistics (including the mean, standard deviation, and range).

The observed-values approach described in this checklist alleviates concerns about extrapolation (because it is based on the observed data) and whether the effect size is representative of the rest of the sample (because it averages over all observed effect sizes). In reality, there are N plausible values, so a complete interpretive strategy must reflect this sensitivity. The sensitivity of the QIs is depicted as a function of the offending variable, and the overall variability is summarized with the standard deviation and range.

Illustrations

In this section I explore three recent examples that demonstrate the drawbacks of the average-case approach and illustrate the benefits of my interpretive approach.

Personalistic Regimes and Nuclear Proliferation

Way and Weeks (2014) address a puzzle in the nuclear proliferation literature; given the historical record of the

[5] More information about the meta-analysis can be found in the supplementary files.

[6] The problem of inconsistency across TDVs appears to be more common with splines than cubic polynomials. The TDVs are consistent in 62.5 percent of the eight uses of polynomials and only 27.3 percent of the eleven uses of splines.

[7] Together, these measures reveal whether counterfactuals are unrepresentative scenarios and potentially model-dependent. See King and Zeng (2006) for a technical description of these measures.

[8] I modify relative risks into changes in predicted probabilities so that they are comparable.

[9] Though the focus is on TDVs, the insights are generalizable to compression in all nonlinear models.

Table 2. Meta-analysis of the sensitivity of QIs across twenty articles

Citation	*Variable*	*TDV (%)*	*Consistent?*	*QI (%)*	*Range*	*Convex hull*	*% Nearby*	*TDV rank*
Gelpi and Feaver (2002)	Percent veteran	>4 (14)	Yes	–0.037 (22)	[–0.505, –0.0002]	No	0	2/11
Meinke, Staton and Wuhs (2006)	Ideological distance	4 (87)	Yes	–0.570 (54)	[–0.820, –0.080]	No	0	1/7
Salehyan and Gleditsch (2006)	Refugees	5 (24)	Yes	+0.007 (71)	[0.0003, 0.049]	No	0	1/5
Fordham and Asal (2007)	Major power democracy	56 (56)	No	+0.029 (63)	[0.008, 0.317]	No	2	2/4
Gleditsch, Salehyan, and Schultz (2008)	Civil war	27 (62)	No	+0.005 (57)	[0.0006, 0.076]	Yes	10	1/9
Salehyan (2008)	Refugees in initiator	28 (62)	No	+0.012 (83)	[0, 0.105]	Yes	19	1/16
Ostby, Nordas, and Rod (2009)	Distance to neighboring conflict	20 (41)	No	–0.032 (37)	[–0.203, –0.0002]	No	1	1/7
Fuhrmann and Kreps (2010)	Violent conflict	11 (55)	No	+0.025 (88)	[0.0003, 0.644]	No	0	1/11
Cunningham, Bakke, and Seymour (2012)	Violent factions	7 (64)	No	+0.082 (59)	[0.004, 0.213]	No	0	1/6
Kleinberg, Robinson, and French (2012)	Trade concentration	20 (52)	No	–0.0005 (7)	[–0.067, –0.0001]	No	0.1	2/11
Owsiak (2012)	Settled borders	24 (64)	Yes	–0.012 (33)	[–0.058, –0.001]	No	0	1/5
Weeks (2012)	Machine	5 (10)	No	–0.013 (1)	[–0.114, –0.0002]	No	7	4/15
Fuhrmann and Sechser (2014)	Deployment	0 (2)	Yes	–0.0005 (44)	[–0.070, –0.0001]	No	11	1/11
Miller (2014)	Dependence score	23 (50)	Yes	–0.002 (42)	[–0.784, –0.0002]	No	11	8/13
Peterson (2015)	Distortion	21 (50)	Yes	+0.00004 (37)	[0.00001, 0.025]	No	51	1/11
Wu (2015)	Labor endowment	1 (30)	Yes	+0.059 (99)	[0, 0.063]	No	33	1/8
Bapat et al. (2016)	GDP per capita	7 (61)	No	+0.074 (53)	[0.010, 0.335]	No	9	6/9
Bapat and Zeigler (2016)	Terrorists	9 (58)	Yes	+0.033 (48)	[0.003, 0.209]	No	7	7/11
Bayer and Urpelainen (2016)	Democracy	9 (50)	No	+0.032 (72)	[0.0004, 0.348]	No	16	2/19
Casper (2017)	Program initiation	3 (9)	Yes	+0.060 (67)	[0.0002, 0.173]	No	2	3/19

Note: Columns list the variable of interest, the value of the TDV in the scenario (and its percentile), whether it is consistent across TDVs, the QI presented in the manuscript (and its percentile), the range of in-sample QIs, whether it is in the convex hull, the percentage of cases that are within one geometric variability ("nearby"), and the rank of *t* relative to the independent variables in generating sensitivity.

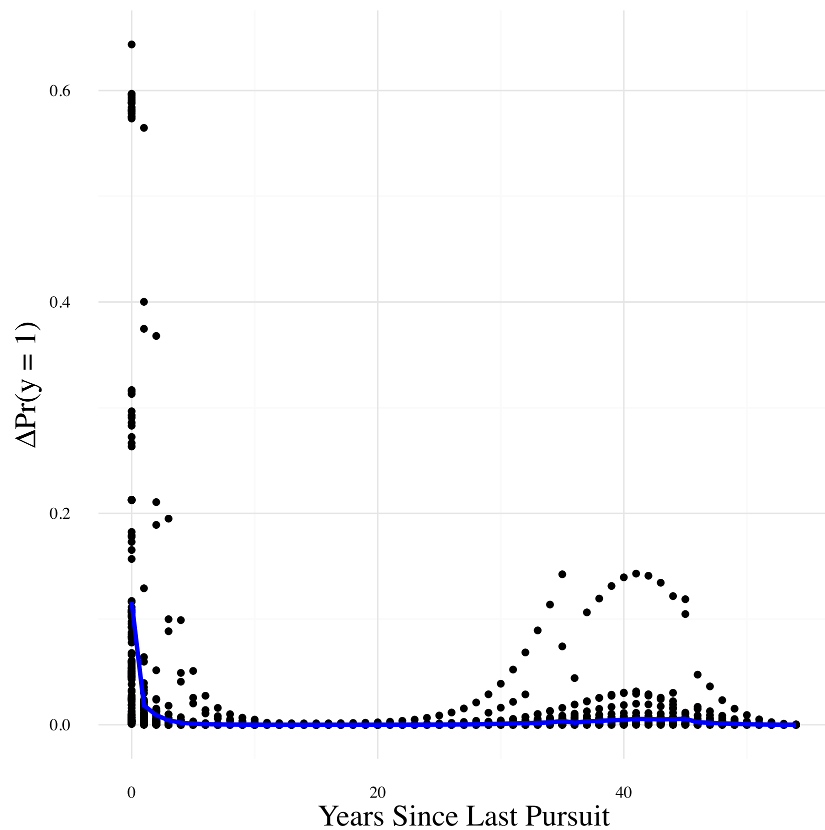

Figure 1. Partial effects of personalist regimes on the probability of nuclear weapons pursuit across values of *years since last pursuit*: Way and Weeks (2014)

Note: The dots represent the first difference of the predicted probabilities (partial effects) of pursuing nuclear weapons for personalist regimes versus all other regimes across the range of values of *years since last pursuit*. The solid line provides the first difference averaged over all the observations (mean partial effects) for that particular value of *years since last pursuit*.

types of regimes that proliferate, why have scholars been unable to discern any relationship between regime type and proliferation? Way and Weeks advocate moving beyond a simple democracy/dichotomy and instead disaggregate dictatorships into various types. They theorize that personalistic regimes are particularly prone to pursue nuclear weapons because they have more motives and face fewer institutional constraints. To control for temporal dependence, they use cubic polynomials measuring the years since the last pursuit of nuclear weapons (Beck et al. 1998; Carter and Signorino 2010). Across a series of model specifications and different codings of proliferation, the authors find strong and consistent evidence that personalistic regimes are more likely to pursue nuclear weapons. The authors, however, miss out on an opportunity to explore the substantive effects of regime type by presenting predicted probabilities or other QIs.

In the section above I provide a checklist for exploring QIs in the context of temporal dependence. The discussion above suggests we should first try to identify the source of the sensitivity. A likely culprit is the group of TDVs, because their large ranges and substantively meaningful coefficients heavily influence an observation's location along the CDF. Figure 1 demonstrates the sensitivity of partial effects as a function of the *years since last pursuit*. The dots represent partial effects of *personalist regimes* calculated for each observation, and the solid line represents the average partial effect at each value of *years since last pursuit*. Interpreting the partial effects in this manner lends additional nuance to the inference that "personalist regimes are more likely than other regime types to pursue nuclear weapons" (Way and Weeks 2014), and it is clear that time shapes this process. The solid line shows that, on average, the changes in probability for a personalist regime will be much larger at low values of years since last pursuit relative to moderate or high values. Indeed, the average partial effect across years since last pursuit ranges from 0.00001 to 0.12, which suggests that the passage of time plays a massive role in determining the propensity for personalist leaders to acquire nuclear weapons.[10]

By considering the distance between the average partial effect (solid line) and the in-sample partial effects (jittered dots), one can see the danger in only selecting one observation to serve as the representative for the entire sample. Unfortunately, this is the exact approach that the vast majority of scholars take with both the average-case and observed-value approaches. They depict one partial effect calculated at either the average value or the mean of the partial effects, while ignoring the thousands of other, arguably more plausible, changes. While the mean partial effect of personalist regimes is 0.007, there is substantial variability (standard deviation of 0.043), and the effect ranges from nearly 0 to a maximum of 0.64. Choosing to present only one partial effect is therefore highly misleading, understates the true variability of the effects, and runs the risk of misleading inferences.

[10] Given that the TDVs capture unmodeled duration dependence, Figure 1 hints at some omitted variable that causes the influence of personalist regimes to wax and wane with time.

Table 3. Representativeness checks for alternative plausible scenarios based on t from the two illustrations that provide QIs

Counterfactuals (t)	*Extrapolate?*	*% Nearby*[a]
Cunningham (2013)		
Replication	No	13
Mean (inconsistent)	No	23
Mean (consistent)	No	26
Median	Yes	35
Mode	No	39
Flores-Macias and Kreps (2013)		
Replication	Yes	0
Mean (inconsistent)	Yes	12
Mean (consistent)	Yes	18
Median	Yes	20
Mode	Yes	21

Note: [a]proportion of observations within one geometric variability (0.18, and 0.15, respectively) of the counterfactual. The other variables are held at their means (continuous) or medians (binary).

Opposition Factions and Civil War Onset

Bargaining models of civil war posit that the inability of nonstate actors to reach an agreement with the state is the result of actors' incentives to misrepresent information and the presence of credible commitment problems (Fearon 1995). A common, and seemingly innocuous, assumption of bargaining models is that the "actors engaged in the bargaining process are unitary" (Cunningham 2013, 660). Cunningham (2013) demonstrates that this assumption is empirically false, which is important because fragmented oppositions have a more difficult time reaching an agreement to prevent civil war. Indeed, in an analysis of all self-determination movements from 1960 to 2005, Cunningham (2013) finds a strong, positive relationship between the number of factions and the risk of civil war onset. This effect is substantively important, as "moving from the minimum to maximum values on the logged [self-determination] factions variable leads to a 37.0 percent increase in the likelihood that a [self-determination] movement will be in a civil war in a given year" (Cunningham 2013, 667). To account for temporal dependence, Cunningham (2013) includes a time counter (*years since civil war onset*) and three cubic splines.

The scenario that Cunningham uses to generate substantive effects (Cunningham 2013) holds "all other variables at their mean, median, or mode" (667n30). I argue that the average-case approach—when coupled with complex and confusing cubic splines—increases the risk of extrapolation. Table 3 shows whether the scenario is an extrapolation (that is, whether it falls outside of the convex hull) and the proportion of observations that are "nearby" (or within 1 geometric variability) (King and Zeng 2006). The scenario that Cunningham (2013) uses to generate QIs is not extrapolated, but it is quite far from the bulk of the observations (only 13.0 percent are nearby).[11]

The dots in Figure 2 represent the in-sample partial effects of an increase in *logged factions* from its minimum (0) to its maximum (3.66), and the solid line represents the mean partial effect calculated at each value of *time since civil war onset*. Beyond the confusion as to what the time variables are actually held to, it is not quite clear what the appropriate average case would look like (for example, Hanmer and Kalkan 2013). When we combine having multiple defensible average scenarios with the sensitivity of the partial effects, the average-case approach raises the potential for serious misrepresentation of the substantive effects. Consider the following three strategies (represented by circles in Figure 2). First, one could offer the mode as the most representative value of the counter variable and therefore calculate the partial effects based on a value of 0 for *years since civil war incidence*. The second strategy would recognize from the histogram at the bottom of Figure 2 that the time counter is most likely right-skewed (skewness = 0.62) and insist that a more representative value would be the median (a value of 10). The final strategy of selecting the mean (a value of 15) is more difficult to justify than the others and is likely the result of the general practice of setting all the independent variables to their means. Table 3 reveals that some plausible average scenarios may be extrapolated values yet close to a large proportion of cases (the median scenario) or interpolated values that are far from the bulk of the data (the replication scenario).

The QIs—and the inferences one would derive—are quite sensitive to this choice. Examining the mean partial effect at the median (10) of years since civil war onset provides a rather modest increase of only 0.10. Compare this to the effect when years since civil war onset is at the mean (15), 0.23, an increase of 133.0 percent over the effect at the median. One of the largest differences occurs when we examine the modal scenario (0), 0.32, an increase of 235.0 percent over the effect at the median. It is important to note that these effects are correct estimates for that particular scenario and are therefore defensible if that scenario is of particular substantive or theoretical interest.

The in-sample partial effects of *logged factions* in Figure 2 reveal how representative the average scenario is to the rest of the sample and how sensitive the effects are to time. Recall Cunningham's (2013) estimate of a 0.37 increase in the probability of a civil war incidence calculated based on the average-case approach. With an average partial effect of 0.20 and a standard deviation of partial effect of 0.14 across the sample, the scenario chosen by Cunningham is certainly at the high end of possible scenarios (eighty-fourth percentile). Figure 2 shows that it is just as reasonable to replace the 0.37 with 0.10 or 0.23 or 0.32. In fact, any other value in the range of partial effects from 0.025 to 0.521 is justifiable based on the data. Moreover, displaying only one value gives it an undeserved status as *the* representative value, when in fact it is only 1 of N possible estimates for that model. The lack of transparency in political science means that scholars have significant leeway in selecting average cases that provide the most meaningful substantive effects. This concern is justified, as the survey described above shows that a surprising number of studies do not describe the simulation scenario, and only 16.9 percent calculate their effects at multiple scenarios. The end result is that readers are often not privy to these decisions, and it is the rare exception when scholars demonstrate the inferential consequences of those decisions.

Partisan Disposition and Implementation of War Taxes

For the final example, I turn to Flores-Macias and Kreps' (2013) analysis of the role of partisanship in determining how American presidents finance war. Out of concern that the "likelihood of the adoption of a war tax at time t may

[11] I believe that this low value is due to two decisions: first, the value of *kin* is set to 0 instead of its median (1), and, second, the treatment of the TDVs is inconsistent. Rather than setting the temporal splines to represent the appropriate transformations of the average value of t, these variables are set to their means (which represent three different values of t).

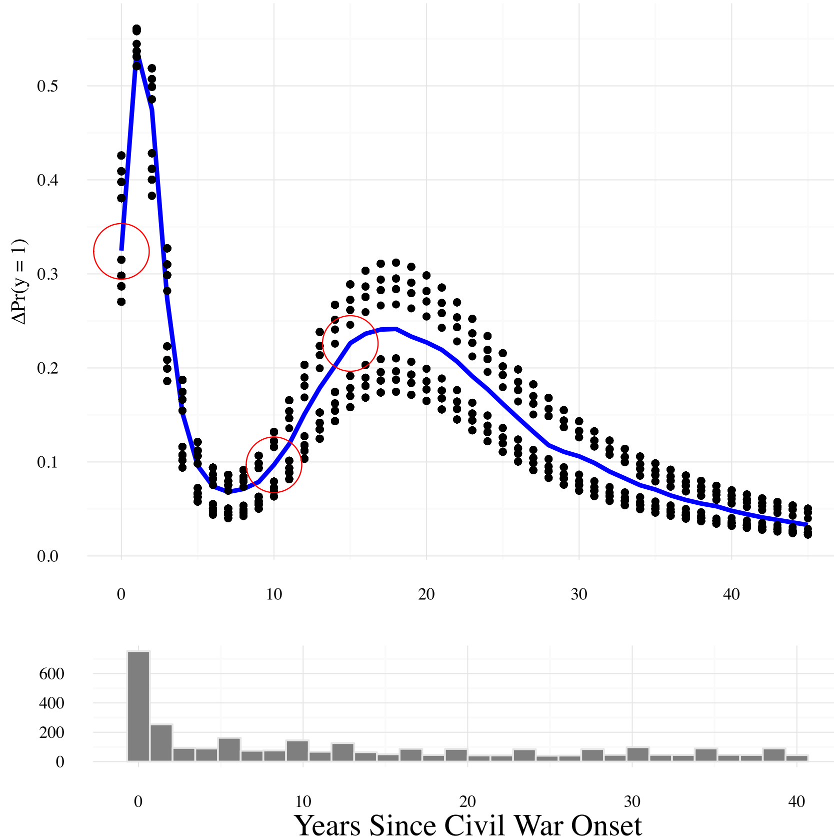

Figure 2. In-sample partial effects of an increase in *logged factions* on civil war incidence across *years since civil war onset*: Cunningham (2013)

Note: Dots represent in-sample partial effects of an increase in *logged factions* from its minimum (0) to its maximum (3.66). The solid line shows the mean partial effect calculated at each value of *years since civil war onset*. Circles depict the mode (0), mean (15), and median (10) values of *years since civil war onset*.

be related to the adoption of a war tax earlier in time," the authors include the *years since war tax* and three cubic splines (Flores-Macias and Kreps 2013, 841). Armed with an original dataset of the instances of war tax adoption from 1789 to 2010, the authors find that presidents from parties with a protax inclination (*party*) are likely to choose taxes to finance the war effort. Based on the most fully specified model (Model 6, Table 2 in Flores-Macias and Kreps 2013), the authors conclude that having a protax president increases the probability of adopting a war tax by 0.061 (Flores-Macias and Kreps 2013). This is based on one simulation scenario in which the other variables are held at their medians (dummy variables) and means (continuous variables).

Since Flores-Macias and Kreps (2013) rely on the average-case approach, it is worthwhile to depict how time influences the size of the partial effects and how the average effect compares to the in-sample partial effects. The dots in Figure 3 represent in-sample partial effects of party on the probability of a war tax, and the solid line represents the mean partial effect at each value of years since war tax. The largest mean partial effect of party occurs at low and high values of years since war tax with smaller mean partial effects at moderate values of t (that is, 24–28). It should also be noted that the TDVs are not statistically significant. One might argue that the time-dependent relationships are properly modeled and can therefore be ignored. Figure 3 shows that this is a risky strategy since the effects of party are highly sensitive to the value of the TDVs, *even though they are not statistically significant.*

Recall that Flores-Macias and Kreps (2013) concluded that having a protax party increased the probability of a war tax by 0.061 (represented by the horizontal dashed line). This effect is at the lower end of the in-sample partial effects (twentieth percentile) and is nearly a full standard deviation (0.15) below the mean partial effect (0.2). This QI drastically understates the influence of presidents' taxation preferences on war funding because of the extrapolated values used in the average-case simulation scenario.[12] All of the continuous variables—including the four temporal dependence variables—are held at their sample means.[13] Yet this ignores the fact that the natural cubic splines are a nonlinear function of the years since war tax variable. Taking the sample means of all four variables is not the same as taking the mean of years since war tax and finding the values of the cubic splines that are associated with that mean value. The result is that the simulation scenario features an average value of years since war tax of 14.4 and mean cubic spline values associated with values of years since war tax of 22, 23, and 25, respectively. Selecting such an unlikely (and impos-

[12] As Table 3 shows, the scenario is an extrapolation where no observations are nearby.

[13] A major culprit is that excellent programs that produce quantities of interest, such as Clarify (Tomz, Wittenberg, and King 2003) and SPost (Long and Freese 2006), have settings that default values to be 0 or the mean (see Hanmer and Kalkan 2013).

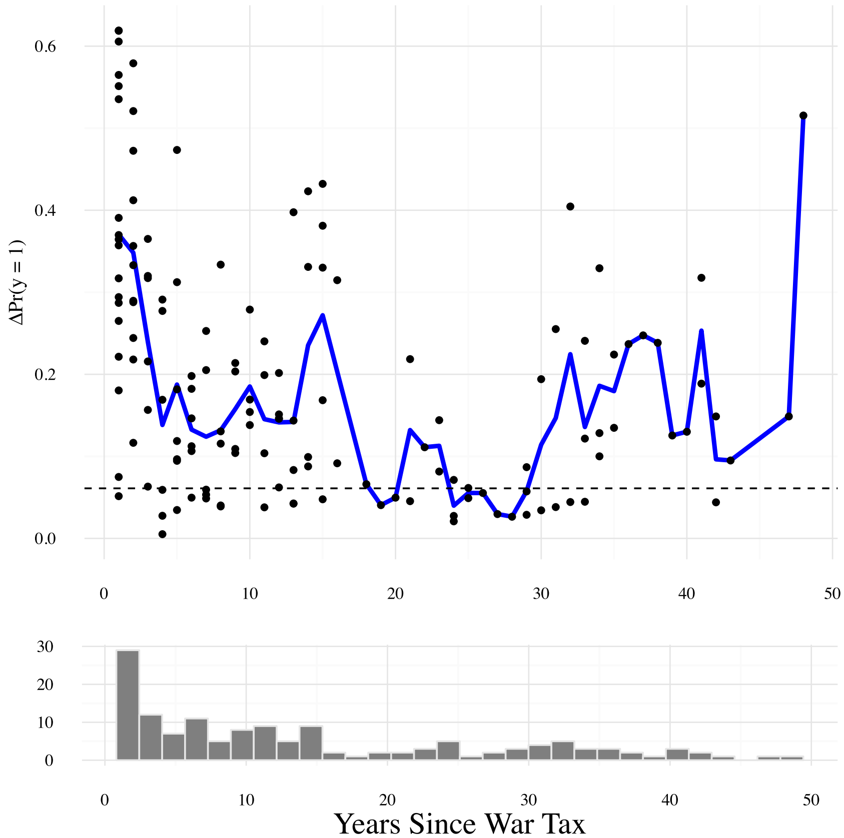

Figure 3. In-sample partial effects of an increase in *party* on war tax across *years since war tax*: Flores-Macias and Kreps (2013)
Note: Dots represent the in-sample partial effects of an increase in *party* from 0 to 1. The solid line shows the mean partial effect calculated at each value of *years since war tax*. The horizontal dashed line represents the partial effect used by Flores-Macias and Kreps (2013) to interpret the results.

sible, given the data) set of values for the simulation scenario causes the partial effect to be much smaller than it should be, by an order of magnitude of one-third to one-fifth.

This inferential error is indicative of the problems associated with using the average-case approach more generally, but especially in models with temporal dependence variables such as splines (which are notoriously difficult to understand) (Carter and Signorino 2010). Setting all of the continuous independent variables to their mean values increases the risk of establishing scenarios that are so unlikely that they drastically under- (for example, Flores-Macias and Kreps 2013) or overstate (for example, Cunningham 2013) the partial effects. Using the observed-value approach avoids this problem since it eliminates the possibility of user error in selecting the correct values of the TDVs. Yet, even the observed-value approach ignores the substantial in-sample variability in partial effects (for example, the dots in Figure 2) because scholars typically only provide the average partial effect. A careful exploration of partial effects therefore requires depicting them across the values of the offending variable (such as t), if necessary, and providing summary statistics of the in-sample partial effects.

Conclusion

Political science models go to great lengths to provide an accurate estimate of the effect of X on Y. As such, scholars spend a significant amount of time trying to get the "correct" model specification. In logit or probit models, this might produce unbiased parameters, but translating these parameters into meaningful QIs is complicated by compression (Rainey 2017). In practice, this means that whether a variable "matters" depends to a large extent on if the simulation scenario selected makes the probability of the event unlikely, equally likely, or likely. I provide evidence that these problems are widespread and they influence the inferences in meaningful ways. With the use of three illustrations from international studies, I demonstrate that compression causes substantial sensitivity in the QIs, problems that are exacerbated by the inclusion of TDVs. Moreover, depicting only one simulation scenario cannot faithfully represent the range of QIs in the sample.

I offer a checklist to guide scholars in their attempts to interpret QIs in the context of temporal dependence. The best solution is to calculate in-sample QIs (which ensures that the scenarios are not extrapolated and are representative of the sample), potentially demonstrate their sensitivity to changes in the covariates' values, and discuss the variability across the sample. International studies would benefit greatly from interpretive methods that display the measures of central tendency and dispersion of the in-sample partial effects and that illustrate how the magnitude changes across time (or other compression-inducing variables). Fortunately, these indicators are easy to calculate, simple to interpret, and significantly improve the accuracy and representativeness of the substantive effects. In the supplementary files I provide detailed code that implements these prescriptions.

The focus of this project is on TDVs, but the inferential problems associated with compression extend far beyond these duration dependence fixes. Scholars must be leery of any control variables that have a substantial influence on the location along the CDF, whether as a result of a wide range, large coefficients, or both. Another set of variables that falls into the same class as TDVs is spatial and temporal fixed effects. Similar to TDVs, they often have meaningful effects (because they soak up a great deal of time- or unit-specific variance), but are often omitted for presentation purposes. The same principles described in this article certainly apply to fixed effects as well. Moreover, this project focuses on logit/probit models, yet the problem of sensitivity arises in nonlinear models generally. The recommendations in this project, therefore, apply similarly to models with count, ordered, or categorical outcomes.

An alternative to the problems described above is to estimate the linear probability model (LPM) via OLS,[14] where the marginal effect of X is the same at all values and regardless of the other covariates' values. Moreover, as Angrist and Pischke (2009, chap. 3) note, the average marginal effect (β) is much easier to recover and is quite close to the average marginal effect from nonlinear models. In the supplementary files I compare the average in-sample marginal effects estimated from the nonlinear model (such as logit or probit) to the marginal effect estimated from the LPM in the three applications and the twenty articles from the meta-analysis. Though the degree of divergence varies widely, in the vast majority of articles (that is, nineteen out of twenty), the LPM marginal effect is quite close (that is, within a standard deviation) to the mean in-sample marginal effect.[15] An obvious drawback of the LPM—and one that is borne out in the survey and meta-analysis—is the likelihood of generating predicted probabilities that fall outside the reasonable bounds of 0 and 1. While the marginal effect is directly interpretable from the OLS coefficients, the ultimate goal—and the one that motivates this project—should be correctly interpreting effects based on the model that is appropriate for the limited nature of the outcome (Long 1997).

Supplementary Information

Supplementary information is available at faculty.missouri.edu/williamslaro and at the *International Studies Quarterly* data archive.

References

Angrist, Joshua D., and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Bapat, Navin, Luis De la Calle, Kaisa H. Hinkkainen, and Elena V. McLean. 2016. "Economic Sanctions, Transnational Terrorism, and the Incentive to Misrepresent." *Journal of Politics* 78 (1): 249–64.

Bapat, Navin, and Sean Zeigler. 2016. "Terrorism, Dynamic Commitment Problems, and Military Conflict." *American Journal of Political Science* 60 (2): 337–51.

Bayer, Patrick, and Johannes Urpelainen. 2016. "It Is All about Political Incentives: Democracy and the Renewable Feed-in Tariff." *Journal of Politics* 78 (2): 603–19.

Beck, Nathaniel, Jonathan Katz, and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42 (4): 1260–88.

[14] I thank an anonymous reviewer for this suggestion.

[15] This is also the case for the three applications with the Way and Weeks (2014) piece being the exception; the LPM marginal effect would be in the ninety-seventh percentile of in-sample partial effects from the nonlinear model.

Berry, William D., Jacqueline H. R. DeMeritt, and Justin Esarey. 2010. "Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential?" *American Journal of Political Science* 54 (1): 248–66.

Carter, David B., and Curtis S. Signorino. 2010. "Back to the Future: Modeling Time Dependence in Binary Data." *Political Analysis* 18 (3): 271–92.

Casper, Brett A. 2017. "IMF Programs and the Risk of a Coup D'etat." *Journal of Conflict Resolution* 61 (5): 964–96.

Cunningham, Kathleen Gallagher. 2013. "Actor Fragmentation and Civil War Bargaining: How Internal Divisions Generate Civil Conflict." *American Journal of Political Science* 57 (3): 659–72.

Cunningham, Kathleen Gallagher, Kristin M. Bakke, and Lee J. M. Seymour. 2012. "Shirts Today, Skins Tomorrow: Dual Contests and the Effects of Fragmentation in Self-Determination Disputes." *Journal of Conflict Resolution* 56 (1): 67–93.

de Boef, Suzanna, and Luke Keele. 2008. "Taking Time Seriously: Dynamic Regression." *American Journal of Political Science* 52 (1): 184–200.

Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49 (3): 379–414.

Fjelde, Hanne, and Nina von Uexkull. 2012. "Climate Triggers: Rainfall Anomalies, Vulnerability, and Communal Conflict in Sub-Saharan Africa." *Political Geography* 31 (7): 444–53.

Flores-Macias, Gustavo A., and Sarah E. Kreps. 2013. "Political Parties At War: A Study of American War Finance, 1789–2010." *American Political Science Review* 107 (4): 833–48.

Fordham, Benjamin O., and Victor Asal. 2007. "Billiard Balls Or Snowflakes? Major Power Prestige and the International Diffusion of Institutions and Practices." *International Studies Quarterly* 51 (1): 31–52.

Fuhrmann, Matthew, and Sarah E. Kreps. 2010. "Targeting Nuclear Programs in War and Peace: A Quantitative Empirical Analysis, 1941–2000." *Journal of Conflict Resolution* 54 (6): 831–59.

Fuhrmann, Matthew, and Todd S. Sechser. 2014. "Nuclear Strategy, Nonproliferation, and the Causes of Foreign Nuclear Deployments." *Journal of Conflict Resolution* 58 (3): 455–80.

Gelpi, Christopher, and Peter D. Feaver. 2002. "Speak Softly and Carry a Big Stick? Veterans in the Political Elite and the American Use of Force." *American Political Science Review* 96 (4): 779–93.

Gleditsch, Kristian Skrede, Idean Salehyan, and Kenneth Schultz. 2008. "Fighting At Home, Fighting Abroad: How Civil Wars Lead to International Disputes." *Journal of Conflict Resolution* 52 (4): 479–506.

Gujarati, Damodar. 2003. *Basic Econometrics*. New York: Mc-Graw Hill.

Hafner-Burton, Emilie M., and Alexander H. Montgomery. 2008. "Power Or Plenty: How Do International Trade Institutions Affect Economic Sanctions?" *Journal of Conflict Resolution* 52 (2): 213–42.

Hanmer, Michael J., and Kerem Ozan Kalkan. 2013. "Behind the Curve: Clarifying the Best Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent Variable Models." *American Journal of Political Science* 57 (1): 263–77.

Harff, Barbara. 2003. "No Lessons Learned from the Holocaust? Assessing Risks of Genocide and Political Mass Murder Since 1955." *American Political Science Review* 97 (1): 57–73.

Keele, Luke. 2008. *Semiparametric Regression for the Social Sciences*. New York: Wiley and Sons.

King, Gary, and Langche Zeng. 2006. "The Dangers of Extreme Counterfactuals." *Political Analysis* 14 (2): 131–59.

Kleinberg, Katja B., Gregory Robinson, and Stewart L. French. 2012. "Trade Concentration and Interstate Conflict." *Journal of Politics* 74 (2): 529–40.

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

Long, J. Scott, and Jeremy Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata*. 2nd ed. College Station, TX: Stata Press.

Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Meinke, Scott R., Jeffrey K. Staton, and Steven T. Wuhs. 2006. "State Delegate Selection Rules for Presidential Nominations, 1972–2000." *Journal of Politics* 68 (1): 180–93.

Miller, Nicholas L. 2014. "The Secret Success of Nonproliferation Sanctions." *International Organization* 68 (4): 913–44.

Milner, Helen V., and Keiko Kubota. 2005. "Why the Move to Free Trade? Democracy and Trade Policy in the Developing Countries." *International Organization* 59 (1): 107–43.

Nagler, Jonathan. 1994. "Scobit: An Alternative Estimator to Logit and Probit." *American Journal of Political Science* 38 (1): 230–55.

Ostby, Gudrun, Ragnhild Nordas, and Jan Ketil Rod. 2009. "Regional Inequalities and Civil Conflict in Sub-Saharan Africa." *International Studies Quarterly* 53 (2): 301–24.

Owsiak, Andrew P. 2012. "Signing Up for Peace: International Boundary Agreements, Democracy, and Militarized Interstate Conflict." *International Studies Quarterly* 56 (1): 51–66.

Peterson, Timothy M. 2015. "Insiders Versus Outsiders: Preferential Trade Agreements, Trade Distortions, and Militarized Conflict." *Journal of Conflict Resolution* 59 (4): 698–727.

Philips, Andrew Q., Amanda Rutherford, and Guy D. Whitten. 2016. "Dynamic Pie: A Strategy for Modeling Trade-Offs in Compositional Variables Over Time." *American Journal of Political Science* 60 (1): 268–83.

Rainey, Carlisle. 2016. "Compression and Conditional Effects." *Political Science Research and Methods* 4 (3): 621–39.

———. 2017. "Transformation-Induced Bias: Unbiased Coefficients Do Not Imply Un- biased Quantities of Interest." *Political Analysis* 25 (3): 402–09.

Salehyan, Idean. 2008. "The Externalities of Civil Strife: Refugees As a Source of International Conflict." *American Journal of Political Science* 52 (4): 787–801.

Salehyan, Idean, and Kristian Skrede Gleditsch. 2006. "Refugees and the Spread of Civil War." *International Organization* 60 (2): 335–66.

Simmons, Beth A. 2000. "International Law and State Behavior: Commitment and Compliance in International Monetary Affairs." *American Political Science Review* 94 (4): 819–35.

Tomz, Michael, Jason Wittenberg, and Gary King. 2003. "Clarify: Software for Interpreting and Presenting Statistical Results." *Journal of Statistical Software* 8 (1): 1–29.

Train, Kenneth E. 2009. *Discrete Choice Methods with Simulation*. 2nd ed. Cambridge: Cambridge University Press.

Von Stein, Jana. 2005. "Do Treaties Constrain Or Screen? Selection Bias and Treaty Compliance." *American Political Science Review* 99 (4): 611–22.

Walter, Barbara F. 2004. "Does Conflict Beget Conflict? Explaining Recurring Civil War." *Journal of Peace Research* 41 (3): 371–88.

Way, Christopher, and Jessica L.P. Weeks. 2014. "Making It Personal: Regime Type and Nuclear Proliferation." *American Journal of Political Science* 58 (3): 705–19.

Weeks, Jessica L. 2012. "Strongmen and Straw Men: Authoritarian Regimes and the Initiation of International Conflict." *American Political Science Review* 106 (2): 326–47.

Williams, Laron K. 2016. "Long-Term Effects for Models with Temporal Dependence." *Political Analysis* 24 (2): 243–62.

Williams, Laron K., and Guy D. Whitten. 2012. "But Wait, There's More! Maximizing Substantive Inferences from TSCS Models." *Journal of Politics* 74 (3): 685–93.

Wu, Wen-Chin. 2015. "When Do Dictators Decide to Liberalize Trade Regimes? Inequality and Trade Openness in Authoritarian Countries." *International Studies Quarterly* 59 (4): 790–801.

Zorn, Christopher. 2000. "Modeling Duration Dependence." *Political Analysis* 8 (3): 367–80.