Advance Access publication March 29, 2016

Political Analysis (2016) 24:243–262
doi:10.1093/pan/mpw009

Long-Term Effects in Models with Temporal Dependence

Laron K. Williams
Department of Political Science, University of Missouri, Columbia, MO, USA
e-mail: williamslaro@missouri.edu

Edited by: Patrick Brandt

A dominant trend in models with binary outcomes is to control for unmodeled duration dependence by including temporal dependence variables. A second, distinct trend is to interpret both the short- and long-term effects of explanatory variables in autoregressive models. While the first trend is nearly ubiquitous in models with binary outcomes, the second trend has yet to be applied consistently beyond models with continuous outcomes. While scholars use temporal splines and cubic polynomials to model the underlying hazard rate, they have neglected the fact that this causes the explanatory variables to have a long-term effect (LTE) by modifying the future values of the temporal dependence variables. In this article, I propose a simple technique that estimates a wide range of probabilistic LTEs in models with temporal dependence variables. These effects can range from simple LTEs for a one-time change in an explanatory variable to more complex scenarios where effects change in magnitude with time and compound across repeated events. I then replicate Clare's (2010, Ideological fractionalization and the international conflict behavior of parliamentary democracies. *International Studies Quarterly* 54:965–87) examination of the influence of government fractionalization on conflict behavior to show that failing to interpret the results within the context of temporal dependence underestimates the total impact of fractionalization by neglecting LTEs.

1 Introduction

In their 1998 seminal piece, Beck, Katz, and Tucker (1998, 1261) stress that "[binary time-series cross-sectional] (BTSCS) data are grouped duration data." As such, one can borrow techniques from event history analysis to appropriately model temporal dependence in binary models (see also Carter and Signorino 2010). While they disagree over the appropriate methodological approaches, they agree on the necessity of incorporating temporal dependence variables in models where the underlying probability of the event varies as a function of time. Scholars of international relations (IR) have largely followed these prescriptions due to concerns that a significant portion of interstate behavior is related to potentially unobservable conditions that correlate with time. It has become nearly ubiquitous that logit or probit models in IR include variables derived from a simple counter of the time since the occurrence of the event, such as dummy variables, cubic polynomials, and cubic splines (see the review in Carter and Signorino 2010, 272). Scholars have used these techniques to answer a wide range of questions in political science; in examining the role of policy diffusion in anti-smoking policies (Shipan and Volden 2008) and death penalty reform (Mooney and Lee 2000); in studies of the effects of oil wealth on regime stability (Smith 2004) and the incidence of party switching in Brazil's Chamber of Deputies (Desposato 2006); compliance with international law (Simmons 2000); and international conflict (Oneal and Russett 1999), among others. Indeed, the techniques described in this article apply to any temporally dependent discrete outcome.

Authors' note: Replication materials are available at http://dx.doi.org/10.7910/DVN/Q7V70C, Harvard Dataverse, V1, February 26, 2016. Supplementary materials for this article are available on the *Political Analysis* Web site (Williams 2016). I would like to thank Patrick Brandt, David J. Brule, David Carter, Justin Esarey, Christopher Gandrud, Mark Nieman, Stephen Quackenbush, Toby Rider, Clint Swift, Vera Troeger, and Zuzana Ringlerova for their invaluable comments. A previous version of this article was presented at the 2013 Annual Meeting of the Midwest Political Science Association and at the University of Texas–Dallas.

Much effort has been made in the past few years to provide a complete interpretation of effects in models with lagged dependent variables in ordinary least squares (OLS) models. Among these projects include suggestions to use autoregressive distributed lag (ADL) models to test the common constraints of lagged and concurrent variables, provide long-term effects (LTEs) (de Boef and Keele 2008), and demonstrate long-run equilibria through dynamic simulations (Williams and Whitten 2012). I suggest that a similar shift in interpretive methods is warranted in models with temporal dependence variables. While temporal dependence fixes have greatly improved the specification of models with categorical outcomes, their deployment has presented a unique problem with respect to the correct interpretation of these models. Put simply, incorporating temporal dependence variables allows the past to influence current values, while at the same time allowing current values to influence the future. While the former point is used to justify the inclusion of splines and cubic polynomials, the latter point has been largely neglected. This is unfortunate, since it means that scholars neglect that explanatory variables may also have an LTE on the outcome of interest. These effects can range from simple LTEs for a one-time change in an explanatory variable to more complex situations where effects change in magnitude with time and compound across repeated events.

Fortunately, calculating LTEs only involves two simple steps. First, calculate the probability of the event occurring at time t. This is the quantity that scholars are most interested in when using logit or probit, but this quantity has another, vastly underappreciated meaning: the probability of an LTE. Second, the LTE is the change in the probability of the outcome at future periods, given that the time variable has been reset to 0 on account of the event occurring at time t. The presentation of probabilistic LTEs uses information that scholars already have at their disposal (such as the baseline probability) and thus requires little additional effort beyond the techniques used to present quantities of interest (and uncertainty). By incorporating the insight of time series models (e.g., de Boef and Keele 2008) and a renewed focus on generating meaningful quantities of interest (King, Tomz, and Wittenberg 2000), these techniques offer a wider range of hypothesis tests regarding long-run dynamics in BTSCS models. I also contribute to the burgeoning debate on the usefulness of temporal dependence variables in testing theories (e.g., Beck, Katz, and Tucker 1998; Carter and Signorino 2010).

I replicate Clare's (2010) piece on the influence of government fractionalization on conflict behavior to demonstrate the inferential consequences of modeling temporal dependence in this manner. Shifts in the explanatory variables change the probability of experiencing an LTE, either shifting the probability down (in the case of negative coefficients) or up (for positive coefficients). The sign and the magnitude of the LTE depend on the strength of the duration dependence and its functional form (i.e., positive, negative, or non-monotonic). By neglecting these probabilistic LTEs, Clare (2010) underestimates the lasting impact of government fractionalization as well as all of the other explanatory variables.

I first situate this piece within the literature advocating the use of temporal dependence variables. I then show how to calculate a wide range of substantively interesting LTEs. LTEs are quantities of interest that are applicable in a number of research situations, as I show in the case of non-proportional hazards and with alternatives to the Beck, Katz, and Tucker approach to modeling duration dependence. In the section that follows, I discuss a series of Monte Carlo experiments that demonstrate the performance of different techniques (ranging from temporal dummies to cubic polynomials and splines) of modeling temporal dependence under varying conditions. I use the insight from these experiments to offer guidelines about the proper use and interpretation of various estimation techniques. Next, I illustrate these concepts in a replication of a project that examines the influence of ideological fractionalization within a government coalition on the probability of dispute initiation (Clare 2010). In the final section, I discuss the applicability of these methods to a general class of models with limited dependent variables.

2 Temporal Dependence

As Beck, Katz, and Tucker (1998, 1261) illustrate, BTSCS data "are grouped duration data." This breakthrough lays the foundation for using event history analysis to control for the possibility that observations will be temporally dependent. BTSCS models estimate a binary dependent variable, y, based on a vector of independent variables, $\boldsymbol{x_{i,t}}$, where observations are indexed by unit i and time period t. If one wants to control for possible temporal dependence, one can explicitly model the hazard rate ($h(t)$) with the following logistic formula (Beck, Katz, and Tucker 1998, 1268):

$$\Pr(y_{i,t} = 1|\boldsymbol{x_{i,t}}) = \frac{1}{1 + e^{-(x_{i,t}\beta + h(t)).}} \tag{1}$$

Event history models offer the promise of a variety of ways of characterizing the functional form of the temporal dependence (such as natural cubic splines) and in dealing with multiple failures (i.e., the number of previous events). A more general specification is the following:

$$\Pr(y_{i,t} = 1|\boldsymbol{x_{i,t}}) = \frac{1}{1 + e^{-(x_{i,t}\beta + t_{i,t}\gamma),}} \tag{2}$$

where t reflects the manner of temporal dependence specification (typically *time since previous event*), whether as dummy variables, cubic polynomials, or some version of splines, and γ characterizes the strength and shape of the influence of t. Beck, Katz, and Tucker's emphasis largely stops there, as they stress that these are simply control variables and should not be treated as theoretical variables (see Beck 2010, 294).

Carter and Signorino (2010) propose cubic polynomials (t, t^2, and t^3) as a more flexible specification of the temporal dependence variables that does not suffer from the problems that plague the usage of temporal dummy variables (such as inefficiency and "quasi-complete separation"). More important for our purposes, Carter and Signorino (280–81, 291) stress that scholars are guided by theory in determining the splines, and then interpret these variables as part of the overall empirical test. This is not an uncontroversial statement, as Beck (2010, 294) claims that "time is not a theoretical variable." More specifically, Beck (2010, 294) claims that the hazard rate "is just a statement about omitted variables [so] duration dependence is both atheoretical and changes with the model."

Of these two camps, I hope to convince scholars of the superiority of the second camp. Whether the source of temporally correlated errors is model under-specification or fundamental uncertainty, scholars include temporal dependence variables—much like unit- or time-specific dummy variables—to account for their particular ignorance about the data-generating process (e.g., Beck, Katz, and Tucker 1998). Yet, there is a great deal of information available in these terms, which can be then used to more completely understand the substantive effects of the theoretical variables. As Bennett (1999, 266) notes, "while duration dependence is indeed theoretically unexplained variance, the pattern of that dependence is informative." The magnitude and the shape of the lasting effects of temporal dependence are closely related to the underlying hazard rate because the effects themselves are a function of changing the values of *time*. I suggest that a full examination of the temporal variables is not only beneficial for these purposes, but required as a means of accessing the entire suite of inferences available from one's empirical model.

While I agree with the practical message of Carter and Signorino, I whole heartedly agree with Beck, Katz, and Tucker's more general point that the goal of scholars in dealing with temporal dependence—as is the case with other "nuisance" parameters such as fixed effects or heteroskedastic parameters—is to provide the "correct" specification that makes these parameters statistically insignificant (Beck 2010, 294; see also Bennett 1999). If the duration dependence is the result of unobserved heterogeneity (or what Zorn [2000, 368] calls "spurious duration dependence"), then one could potentially better specify the model to the point where there is no duration dependence.[1] The simple hypothesis test would be that $\gamma = 0$ in equation (2), and failing to reject this null would support the conclusion of temporal independence. If these coefficients are jointly equal to 0, then the grouped duration model would collapse to an exponential model with a flat underlying hazard (Box-Steffensmeier and Jones 2004, 22).

[1]Zorn (2000: 368–69) differentiates this type from state dependence, or "true duration dependence," which reveals "the propensity of a state toward self-perpetuation."

However, we have seen that even in the case of substantive topics where there is a rich theoretical literature from which to draw (such as international conflict), scholars have been unable to completely specify a model such that the temporal dependence variables do not add explanatory power.[2] For a variety of methodological reasons ranging from measurement error to idiosyncracies, incorrect functional form to fundamental uncertainty, eliminating the temporal dependence by including only conceptual variables may seem like an unattainable goal.

Based on this reality, I suggest that scholars fully interpret their temporal dependence variables regardless of whether or not the hazard rate is itself theoretically interesting. Interpreting the hazard rate is closely related to producing accurate inferences about the long-term substantive effects of the theoretical variables. Modifying the current values of the variables changes the probability that future values of the temporal dependence variables will reset to 0. In much the same way as an OLS model with a lagged dependent variable (de Boef and Keele 2008), the theoretical variables contain both a short-term and an LTE. This probabilistic LTE can reflect a number of different quantities of interest (ranging from one-period changes to permanent shocks), the possibility of compounded effects, and can easily be modified to reflect non-proportional hazards. A thorough interpretation of the hazard rate is, therefore, critically important to making a complete and accurate set of inferences.

LTEs in Models with Temporal Dependence

Given that IR scholars employ temporal dependence variables as a simple means of incorporating country- and dyad-specific histories, it is surprising that scholars treat all the explanatory variables as if their effects occur immediately and without lag. To assess whether these patterns are consistent across the discipline, I surveyed each article published in the *American Political Science Review*, *American Journal of Political Science*, *Journal of Politics*, and *International Studies Quarterly* that cited Beck, Katz, and Tucker (1998). Out of 178 articles, 131 employ some version of temporal dependence variables. Of these, only two discuss the possibility of lasting effects of the independent variables (Bennett 2006; Beardsley 2008), but neither calculates the actual LTEs. It is clear that the vast majority of political science research in the most prestigious journals treats key variables as having an impact that completely occurs at time t with no lasting effects.

This is surprising since scholars often justify the inclusion of temporal dependence variables because "the probability of dyadic conflict in a given year, for example, is likely to be dependent on the conflict history of that dyad" (Beck, Katz, and Tucker 1998, 1263). Studies examining reputations (Crescenzi 2007), "repetitive military challenges" (Grieco 2001), and enduring rivalries (Thompson 2001) posit that international conflict can have lasting ramifications, and this logic extends to other notable phenomena, such as coups (Londregan and Poole 1990) and civil wars (Collier et al. 2003), to name a few. It is, therefore, unreasonable to expect that the effects of the variables in one period do not bleed over into the next period. Other approaches incorporate history by including the lagged dependent variable (or latent variable) on the right-hand side (Jackman 2000). Though intuitively pleasing, transitional models such as these have not caught on to the extent that the Beck, Katz, and Tucker techniques have (for an exception, see Przeworski et al. 2000, 137–39). Nevertheless, it should be noted that any discrete choice model that includes history on the right-hand side—whether it is as Beck, Katz, and Tucker advocate or something more complex like a parameter-driven transitional model—causes the full effects of explanatory variables to take more than one period to complete.

A more complete picture, therefore, treats the coefficients as only short-term effects of X on Y^* and evaluates the LTEs of those variables in the context of the temporal variables. This procedure is complicated slightly because non-linear model specifications do not typically include an observed lagged dependent variable (such as in OLS; see Esarey and DeMeritt 2014).[3] With the inclusion of any specification of time since the previous event (*time*), each change in the explanatory variables

[2]A notable exception is the work by Bennett and Stam (1996), who show that increasingly better specified models of war duration change the temporal pattern from one of dependence to independence.
[3]Transitional models are a notable exception (Jackman 2000).

has an LTE that is characterized probabilistically. In other words, modifying the values of any of the independent variables at time t potentially influences the predicted probabilities of the outcome in future time periods by forcing *time since previous event* to revert back to 0, which itself affects the probability of observing the event.

The first step in calculating LTE is to calculate the predicted probability of the outcome, given a particular configuration of values of the independent variables (or simulation scenario, $\mathbf{X_C}$), $\Pr(\hat{y} = 1|\mathbf{X_C})$. A quantity of interest that is often directly related to the hypotheses is the change in predicted probability given a change in the variable of interest, X_K, or $\Delta\Pr(\hat{y} = 1|\mathbf{X_C}, \Delta X_K)$. Depending on one's theory, a variety of quantities of interest related to long-run dynamics are available, including temporary and permanent changes to an explanatory variable. An LTE occurs if the observed outcome, $\hat{y} = 1|X_{C,}$ changes as a result of the change in X_K. However, since this exercise involves a counterfactual (King and Zeng 2006), we never actually observe $\hat{y}$, only $\Pr(\hat{y} = 1)$.

This is an important distinction, since it determines whether scholars are merely "classifying" observations (in-sample) or "predicting" observations (out-of-sample). For example, consider the common practice of assessing model fit by classifying predicted probabilities into outcomes. Since the actual values of the outcomes are observed for in-sample observations, this practice becomes one of deciding a threshold, π, and classifying those observations with probabilities exceeding that threshold as having a value of 1, and those who do not pass the threshold as having a value of 0 (Greene 2008). Though typically set at $\pi = 0.5$, the threshold for classification is arbitrary and can be modified so that one might increase the classification rate for low-probability events (Greene 2008). Of course, modifying this threshold imposes trade-offs between correctly classifying events and incorrectly classifying non-events, hence one must determine the threshold *a priori* based on the costs associated with each type of classification error (King and Zeng 2006, 632–33).

On the other hand, when dealing with counterfactuals (or other out-of-sample predictions), it makes little sense to classify probabilities into outcomes. Scholars have centered their criticism on model fit statistics such as "percent correctly predicted" and classification tables because it conflates a best guess of $\hat{y}_i$ with actually observing y_i (Herron 1999; Train 2009).[4] Just because a probability exceeds a certain threshold does not imply the occurrence of the event, and making that inference "misrepresents choice probabilities" (Train 2009, 69). For example, a predicted probability of 0.6 for a given simulation scenario means that there is a 0.6 probability of the outcome occurring, or that out of 1000 times, the event will occur 600 times. Extending this logic further into the realm of forecasts, Pindyck and Rubinfeld (1991, 268–69) note that the best one can do is to provide a forecast of the probability of the event occurring, which will *never* be correct *ex post* (since the outcome is either 0 or 1).

Therefore, the likelihood of a variable having an LTE is simply the change in the probability of observing the outcome, $\Delta\Pr(\hat{y} = 1|\mathbf{X_C}, \Delta X_K)$. In addition to reflecting the change in probability of observing the outcome at time t, this quantity of interest characterizes the probability that *time* is reset to 0 at time $t+1$, therefore changing the simulation scenario for *time* at subsequent time periods.[5] Probabilistic LTEs can, therefore, be structured for virtually any kind of theoretically meaningful scenario. For example, the process is flexible enough that one can examine a change in the explanatory variable (ΔX_K) for one time period, a temporary change, or even a permanent change. In this way, the utility of calculating LTEs reflects the variety of quantities of interest to scholars who study long-run dynamics (e.g., de Boef and Keele 2008; Williams and Whitten 2012).

Since the probability of an LTE is based on parameter estimates which are themselves uncertain (King, Tomz, and Wittenberg 2000), it is necessary to characterize this uncertainty with either standard errors or confidence intervals. Fortunately, the calculation of this probability requires

[4]A superior indicator of model fit for in-sample observations is the "expected percent correctly predicted" (ePCP), because it distinguishes between low and high probabilities and comes with measures of uncertainty (Herron 1999).

[5]Alternatively, if one is comfortable using probability thresholds, then for the purposes of these dynamic simulations one can claim that a scenario has an LTE if it increases (decreases) the probability above (below) a cutoff specified *a priori*. For example, scholars using the conventional threshold of $\pi = 0.5$ would conclude that any variable that moves the probability across that threshold has an LTE. A simulation scenario that fails to cross the threshold would have no LTE.

no additional assumptions beyond those used when calculating other quantities of interest.[6] Nonetheless, using the parameter estimates to derive an additional set of inferences warrants some careful consideration of model fit. I encourage scholars to be transparent by providing measures such as ePCP (Herron 1999), cross-validation (Beck 2001), receiver-operating characteristic (ROC) curves (King and Zeng 2006), area under the curve (AUC) (Weidmann and Ward 2010), and separation plots (Greenhill, Ward, and Sacks 2011). Scholars should also conduct a series of diagnostic tests to ensure the appropriate specification of the temporal dependence prior to calculating LTEs (Keele 2008, 109–35).

Once we have characterized the probability of an LTE, the next step is to calculate the LTE. Recall that the LTE is the change in predicted probability of the outcome at periods $t+1$ to $t+k$ in the baseline scenario ($\mathbf{X_C}$) compared to the scenario where the time since previous event (*time*) reverts to 0 at time $t+1$. Since *time* is a counter based on previous periods' values, there is an LTE at each value of *time*, from $t+1$ to $t+k$. In other words, the LTE is a sequence of moving differences in the probability for two points along the hazard rate: one that assumes the event occurred at time t and one that does not. Assume that we set up a baseline scenario, $\mathbf{X_C}$, which represents the average value of the independent variables, including *time* ($\bar{t}$). We then assume that the event occurs at time t so that the otherwise identical simulation scenario now has a value of $time = 0$ at time $t+1$. The LTE at time $t+1$ is the following:

$$\text{LTE}^{t+1}_{\mathbf{X_C}} = \Pr(\hat{y} = 1|\mathbf{X_C}, time = 0) - \Pr(\hat{y} = 1|\mathbf{X_C}, time = \bar{t}). \quad (3)$$

I calculate the LTE at time $t+2$ by updating the values of *time* in both scenarios:

$$\text{LTE}^{t+2}_{\mathbf{X_C}} = \Pr(\hat{y} = 1|\mathbf{X_C}, time = 1) - \Pr(\hat{y} = 1|\mathbf{X_C}, time = \bar{t} + 1). \quad (4)$$

And so on, up to a value of k (representing the maximum or some other intuitive value of *time*). It is important to note that *time*—in addition to all the other temporal dependence variables derived from *time* (such as splines or cubic polynomials)—must be updated at each time period. Fig. 1a–d illustrate four configurations of LTEs depending on one's quantities of interest.[7] These data exhibit a pattern of negative duration dependence, where the probability of the event is highest immediately following its occurrence, and the probability decreases with time.[8]

First consider Fig. 1a as we move from left to right. The first dot (and vertical dashed line) represents the baseline probability (and 95% confidence interval) of the event occurring given the simulation scenario (or $\Pr(\hat{y} = 1|\mathbf{X_C})$). The second dot represents the updated probability of the event occurring given a one-time change in the variable of interest (or $\Pr(\hat{y} = 1|\mathbf{X_C}, \Delta X_K)$). In this case, the probability of the event occurring at time t is about 0.09. The number labels represent the values of t in both scenarios. One can assess whether the change in X_K produces a statistically significant change in the probability of an LTE by determining whether the confidence intervals overlap. In this case, the increase in X_K does not produce a statistically significant change in the probability *for that time period*. The remainder of Fig. 1a, however, reveals that the change in X_K has a meaningful impact on the probability *in future periods* by changing the probability that the t variable resets to 0. The dashed lines from $t+1$ to $t+12$ are the 95% confidence intervals for the probability of the event, given that the event did not occur at time t ($\Pr(\hat{y} = 1|\mathbf{X_C}, time = 14\ldots25)$. The solid lines, on the other hand, represent the counterfactual where $\hat{y}_t = 1$. The two vertical lines at time $t+1$ illustrate how the value of t either resets to 0 (if $Y_t = 1$) or continues beyond its current value (if $Y_t = 0$). Of the two scenarios, the counterfactual where the event does not occur ($Y_t = 0$) is much more likely given its small probability (0.09). The difference between these two vertical lines is the visual representation of the LTE from equation (3). We can also use the confidence intervals to

[6]See King, Tomz, and Wittenberg (2000, 351) for a discussion of the assumptions related to the sampling distribution and model specification.

[7]Replication materials are available from the Harvard Dataverse (Williams 2016).

[8]I generate the data to mimic the typical features of a study examining international conflict. There are 1000 observations, and the negative hazard rate is based on Weibull distribution. The X is drawn from a uniform distribution from -2 to 2, and the substantive effect is for a 0.5 increase in X ($\beta_{X_K} = 1$). The outcome occurs in 17% of the sample, and the t variable is right-skewed (skewness = 1.9) with a mean of 9.

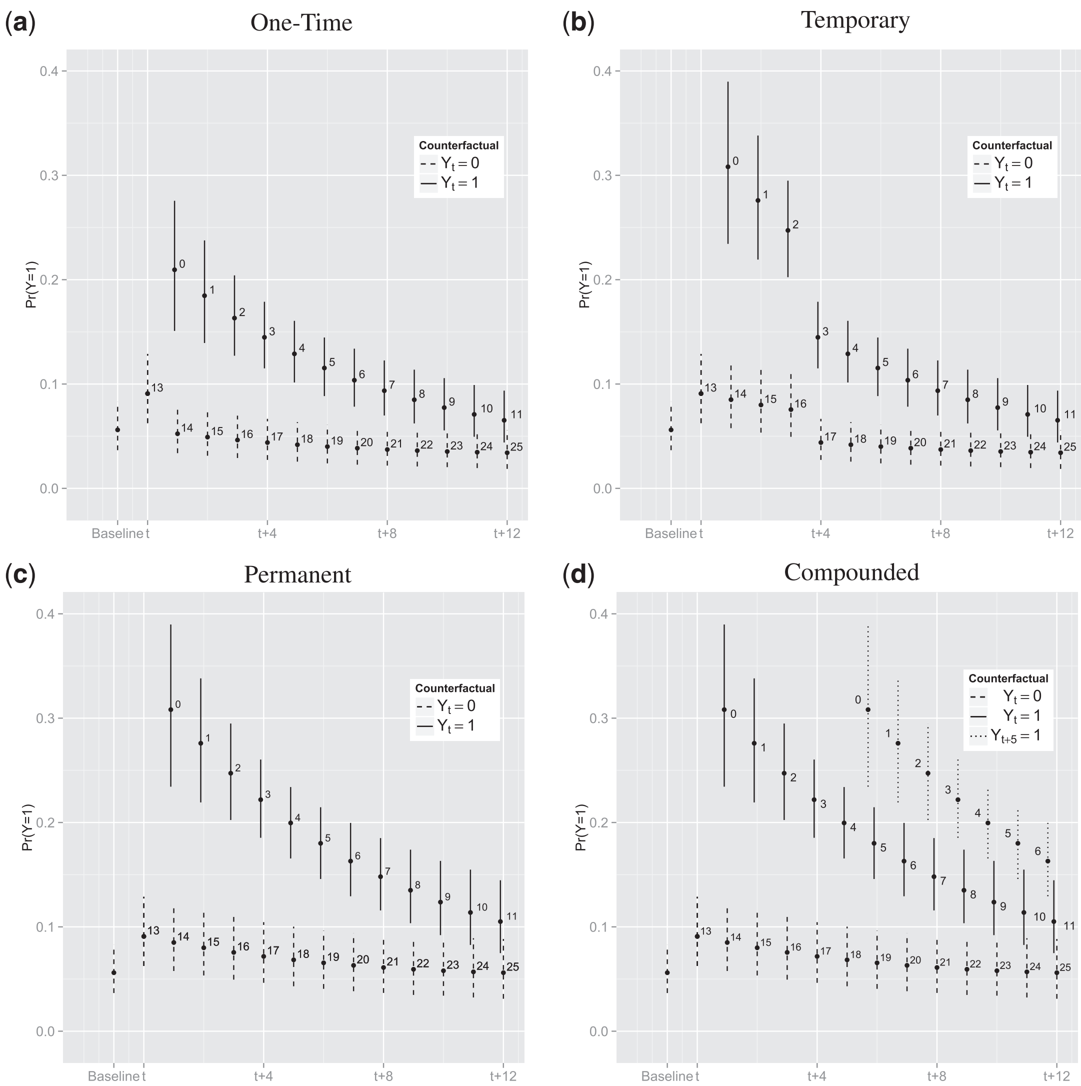

Fig. 1 Variations of LTEs based on the counterfactual of interest for negative duration dependence. *Note:* Numbers attached to points represent the values of t in the scenarios.

conclude that there is a statistically significant LTE from $t+1$ to $t+9$, at which point the confidence intervals overlap and there is no statistical difference between the two probabilities.

One might also be interested in the LTEs of a more lasting increase in X_K. Figure 1b is a slight variation on Fig. 1a in that it depicts the same simulation scenario as before, except that the quantity of interest is the change in long-term probability given an increase in X_K for three periods beyond t. The inferences regarding LTEs are unchanged, and the only difference is that the probability of the event occurring for the two counterfactuals remains higher for those three periods. If one's emphasis is centered on a more lasting effect, then one can set up the simulation scenario to represent a permanent increase in X_K as in Fig. 1c. Again, the inferences remain the same although the probabilities of the two scenarios are consistently higher. It is also easy to infer what other LTEs might look like with different shapes of the hazard rates.[9] A figure depicting the

[9]In the Online Appendix, I provide examples of LTEs from a variety of estimation techniques with different shapes of the underlying hazard rate.

LTE for a positive hazard rate, for example, would rearrange the two counterfactual scenarios so that the scenario where an event occurred at time t would be lower than the other counterfactual. The resulting LTE would be negative.

Up until now, I have simplified the calculation of LTEs by suggesting that the event only happens at time t, and the time variable only reverts back to 0 one time. In reality, increasing the probability of the event in one time period—when combined with negative duration dependence—makes it more likely that the event occurs at future time periods as well. A perfect example of these compounded LTEs is the "conflict trap," or the idea that once a country has experienced a civil war it is much more likely to also have violence in the future. Consider the scenario where a massive decrease in economic development triggers a civil war at time t; the civil war "weakens the economy and leaves a legacy of atrocities. It also creates leaders and organizations that have invested in skills and equipment that are only useful for violence" (Collier et al. 2003, 4). These circumstances make it likely that the initial effects of economic shocks are compounded through multiple civil conflicts. Therefore, a more realistic depiction allows the time counter to reset at multiple time points.

As opposed to the other panels, Fig. 1d depicts two counterfactuals where the event occurs: the solid line represents the resulting probabilities after we assume the event occurs at time t, and the dotted line represents the resulting probabilities after we assume the event occurs at time $t+5$.[10] In addition to the LTE given the event at time t (with probability 0.09), there is a compounding effect that results in an even larger LTE as a result of the event at time $t+5$ (with probability 0.20). This is certainly consistent with our theoretical understanding of a wide range of political phenomena, and it suggests that our empirical estimates of the effects of key theoretical variables are incorrect because scholars have not calculated probabilistic LTEs.

The calculation and effective interpretation of LTE are complicated somewhat because the substantive magnitude of the LTE depends on the values of $\mathbf{X_C}$. Econometricians have long noted that the effects of a variable on the probability of an outcome in a logit or probit model are conditional on the location along the cumulative density function (CDF) (e.g., Nagler 1991; Nagler 1994; Long 1997, 73), which means that the marginal effects are different than those from an OLS model (e.g., Ai and Norton 2003; Norton, Wang and Ai 2004). Though Berry, DeMeritt and Esarey (2010) have recently highlighted the severity of the problem for political scientists, it is clear from the survey that the discipline has yet to fully implement these suggestions.[11] In the context of this project, this problem becomes magnified when one employs temporal dependence variables, because these variables have a wide range and are typically of a high magnitude relative to the other variables.[12]

Consider the formulas used to calculate the LTEs described above. Since the substantive effect of each variable in a logit or probit model is influenced by the $\mathbf{X}\beta$ (or the latent Y^* variable), scholars must consider the LTEs under different simulation scenarios ($\mathbf{X_C}$). Under these circumstances, it is often incomplete, and in other times highly misleading, to interpret the LTEs holding the variables in $\mathbf{X_C}$ at their means or some other arbitrary values (i.e., the so-called "average-case" approach; see Hanmer and Kalkan 2012). This is due to two concerns: first, "mean" values of the variables may not reflect in-sample observations, but instead nonsensical values (King and Zeng 2006). Second, these "mean" values may exaggerate (or minimize) the LTEs by producing an $\mathbf{X}\beta$ value close to 0. One must, therefore, be cautious when choosing appropriate cases, by first choosing values of the simulation scenario that reflect meaningful in-sample observations, and second interpreting the LTEs at multiple, substantively interesting scenarios. Alternatively, one could calculate the

[10]As with the choice of whether or not there is an LTE, the decision to depict compounded LTEs (in this case, at time $t+5$) can arise endogenously (due to the probability passing some threshold), or exogenously by the researcher (to reflect an interesting or historical case). In either case, one should emphasize the probability of the LTE so that the compounded LTEs depicted are not unreasonable or unlikely.

[11]Only 4.7% (6 out of 129) of the articles in the above survey calculated quantities of interest at multiple simulation scenarios to demonstrate the effects of compression.

[12]In the Online Appendix I demonstrate that there is the potential for an observation's location along the CDF (and thus, the size of the substantive effect) to be largely determined by the values chosen for the temporal dependence variables in the simulation scenario.

average LTE using the actual values for each observation in the sample (i.e., the so-called "observed-values" approach; see Hanmer and Kalkan 2013).

LTEs in models with temporal dependence are an important—and largely ignored—quantity of interest in guiding substantive inferences. Much like other quantities of interest, effective presentation of the LTE requires providing both the probability of an LTE and the LTE itself (with measures of uncertainty). Fortunately, the probability of an LTE is already a quantity of interest that scholars typically report, whether as the baseline probability or the change in the probability of an outcome, conditional on a simulation scenario. Since the hyper-conditionality causes the LTE to vary across values of *time* (as well as the other variables), the most comprehensive presentation is to depict the relationship graphically across multiple scenarios. If one's theoretical focus is motivated by particular historical cases, then one can calculate the LTE for that configuration of conditions (i.e., independent variables). Or, if one is more concerned about the "average" LTE, one can calculate the LTE for each in-sample observation and then average over those values (Train 2009; Hanmer and Kalkan 2013). Graphically depicting the LTE (with appropriate measures of uncertainty) offers two intuitive hypothesis tests. First, the null hypothesis that the simulation scenario has no LTE is rejected if the confidence intervals for the LTE do not overlap zero. Second, the null hypothesis that the LTE does not vary across values of *time* is rejected if the confidence intervals do not overlap at distinct values of *time*.

Non-Proportional Hazards

A prominent issue in event history analysis is that most models (such as those described above) assume proportional hazards (Box-Steffensmeier and Zorn 2001; Box-Steffensmeier and Jones 2004; Licht 2011), or that "the effects of covariates are constant over time; the effect of an independent variable is to shift the hazard by a factor of proportionality, and the size of that fact remains the same irrespective of when it occurs" (Box-Steffensmeier and Zorn 2001, 973). Whenever the outcomes are influenced by "processes of learning, institutionalization, strategic developments, and information transmission," the assumption is likely to fail (Licht 2011, 228). For example, in the context of international conflict, it certainly is reasonable to believe that the factors exacerbating the risk of conflict should be most destructive immediately following the previous conflict. Likewise, the constraining effects of conditions are likely to be most influential immediately following a conflict when the risk of repeated violence is high. Since the models force the proportional hazards assumption, we potentially risk systematic bias (Box-Steffensmeier and Jones 2004, 132) when we fail to explicitly model the non-proportional hazards (NPH). In equation (2), modeling non-proportional hazards would entail interacting the t variable(s) with the offending X variable, which would allow the effects of X to vary with elapsed time.

Figures 2a and b depict the probabilities of an event for two counterfactuals ($Y_t = 0$ and $Y_t = 1$) following a permanent increase (from 0 to 0.5) in X_K (from $t+1$ to $t+16$) for a scenario with negative duration dependence and positive non-proportional hazards.[13] Essentially, while the baseline probability of the event declines as a function of elapsed time, the effects of X_K increase with time. Figure 2a is based on an additive specification (based on one time counter, t) that ignores the non-proportional hazards, while Fig. 2b explicitly models the non-proportional hazards with an interaction between X_K and t. The model that ignores the positive NPH estimates a smaller negative hazard rate ($\beta_t = -0.08$), which is reflected in the smaller change in probabilities from $t+1$ to $t+16$. The other model explicitly allows the effects of X_K to increase with t, hence the hazard rate is more negative ($\beta_t = -0.19$), yet this decreasing effect is counterbalanced at high values of t by the positive NPH ($\beta_{Xt} = 0.07$).

Recall that the differences in probabilities in the two counterfactuals (from $t+1$ to $t+16$) form the LTEs. Figure 2c demonstrates the difference in LTEs when modeled (Fig. 2b) compared to

[13]The data are created in a manner similar to those above, with the exception that the influence of X_K now varies as a function of t ($X_K \times t$), and the influence increases with time ($\beta_{Xt} = 0.04$).

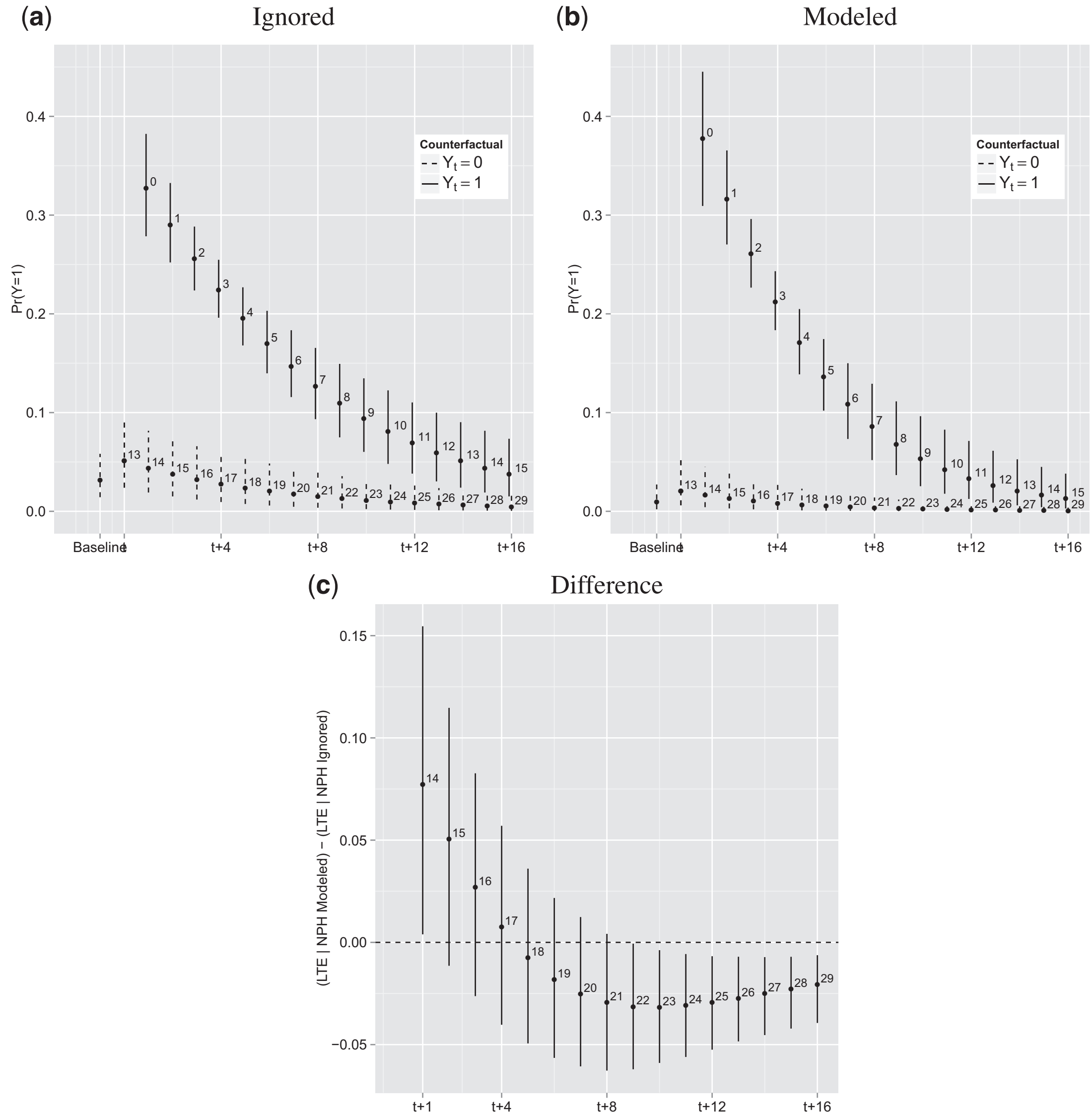

Fig. 2 LTEs with negative duration dependence and positive non-proportional hazards: ignored (left), modeled (right) and the differences in LTEs (bottom).
Note: Numbers attached to points represent the values of t in the scenarios.

ignored (Fig. 2a). When properly modeled, the LTE is statistically higher at early values of the simulation and statistically lower at the later values of the simulation compared to the model that incorrectly assumes proportional hazards. It is clear from this illustrative case that ignoring non-proportional hazards incorrectly estimates the shape of the temporal dependence (β_t) and completely misses the time-varying effects of X_K (Carter and Signorino 2010, 289). Scholars should think carefully about whether the influence of variables will wax or wane with the passage of time and rule out those possibilities with explicit empirical tests prior to calculating LTE (Carter and Signorino 2010, 289).

Alternatives to Duration Dependence

A number of alternatives to dealing with temporal dependence grow out of the time series tradition rather than event history models. *Transitional models* integrate the history of the unit but do so in a way that reflects whether an event occurred in the previous period (*state dependence*) rather than the time since the previous event (*duration dependence*) (Jackman 2000, 6).[14] The interpretation of non-proportional hazards reflects this subtle, but important, difference; instead of an explanatory variable's influence changing with time, its influence varies according to whether the event occurred in the past (such as the previous period in an AR(1) process). Transitional models can either be observation-driven (where the actual observed outcome in the previous period conditions the effects of the covariates) or parameter-driven (where the latent y^*_{t-1} appears as a covariate).

Given the elevated status of history, it is no surprise that LTEs are meaningful quantities of interest in these models. The calculation of LTEs in the observation-driven transition model is similar to the Beck, Katz, and Tucker approach in that one still uses probabilities to make counterfactual claims about whether an event occurs, except that the probabilities determine the previous state (and the values of the interaction terms). The LTEs in the parameter-driven transition model are close analogues to those derived from ADL models in the time series tradition (Jackman 2000, 22–24; de Boef and Keele 2008; Esarey and DeMeritt 2014). Whereas the observation-driven model is much easier to estimate (see Jackman 2000, 10), the parameter-driven model is a closer match to theories that often make claims about serial dependence relating from circumstances that are unobservable. Regardless of the method one chooses to account for temporal dependence, each method offers different strategies for calculating LTEs and for addressing possible non-proportional hazards.

Monte Carlo Experiments

The accurate calculation and effective depiction of LTEs—much like other quantities of interest—"assume that the statistical model is identified and correctly specified" (King, Tomz, and Wittenberg 2000, 351).[15] More specifically, correct estimation of LTE depends on both the estimate of the coefficient of interest (β) and the hazard rate (β_t). The size of the β influences the change in probabilities in the counterfactual scenario; an unbiased β will give an accurate estimate, on average, of the increased (decreased) probability of experiencing the event at future periods. The hazard rate also plays a key role in determining the size of the LTE; the magnitude and sign depend on the value of t in the counterfactual scenario (which is then reset to 0), and the shape of the functional form. It is possible to experience LTEs that are positive, negative, and both, depending on whether shape of the hazard monotonically decreases, increases, or has a non-monotonic shape, respectively. In the Online Appendix, I demonstrate how the inferences regarding LTEs vary based on the functional form of the hazard rate.

Therefore, any discussion of the accuracy of LTE must address the estimate of β and the hazard rate. It is not yet clear how inaccurately modeling temporal dependence influences the calculation of LTE. My goal in this section is to assess the performance of a variety of estimation techniques to model temporal dependence under different conditions. I generate a variety of data sets ranging in size ($N \in \{1000, 5000, 10000\}$)[16] with the binary dependent variable based on an explanatory variable (X_K) drawn from a uniform distribution from -2 to 2 ($\beta_{X_K} = 1$), a negative constant ($\beta_C \in \{-3, -2\}$, to produce a variety of frequencies of events) and a hazard rate with the logistic

[14]Studies of conflict initiation unknowingly make assumptions about non-proportional hazards based on state dependence when coding ongoing years of conflict as 0 in their empirical analyses (Bennett and Stam 2000, 661–62). Unless one estimates a model that allows for state dependence (such as an observation-driven transitional model [see Jackman 2000] or a dynamic probit [see Przeworski and Vreeland 2002]), this decision treats instances of peace and ongoing conflict as the same and forces the effects of independent variables on both processes to be identical (McGrath 2015).

[15]Of course, this focus on "coefficient-induced bias" leaves out "transformation-induced bias'" which arises from the transformation of coefficients into quantities of interest (Rainey 2015). Rainey (2015) demonstrates that quantities of interest do not inherit the small sample properties of unbiasedness that coefficients do. The problem of transformation-induced bias is complicated by the fact that it is tough to characterize the bias generally, especially if it is non-linear (as in the case of logit or probit). Moreover, small sample bias is a bigger concern for quantities of interest than coefficients (Rainey 2015). The remainder of this manuscript focuses on possible problems related to model misspecification (coefficient-induced bias) and briefly discusses transformation-induced bias under the guise of the consequences of compression.

[16]Rainey (2015, 8) demonstrates that transformation-induced bias tends to approach 0 as sample size increases, so I minimize this problem by creating large sample sizes.

Table 1 Performance of β_x under various circumstances in Monte Carlo experiments: negative duration dependence

Scenario	*Avg.* β_{X_K}	*Bias*	*MSE*	*SE*	*SD*
Exponential (Flat)					
$N = 1000$; 1s = 16%	0.964	0.082	0.011	0.098	0.096
$N = 1000$; 1s = 37%	0.964	0.062	0.006	0.073	0.069
$N = 5000$; 1s = 16%	0.962	0.047	0.003	0.044	0.043
$N = 5000$; 1s = 36%	0.959	0.044	0.003	0.033	0.032
$N = 10{,}000$; 1s = 16%	0.959	0.043	0.003	0.031	0.031
$N = 10{,}000$; 1s = 36%	0.959	0.042	0.02	0.023	0.024
Temporal dummies					
$N = 1000$; 1s = 16%	1.002	0.078	0.010	0.101	0.100
$N = 1000$; 1s = 37%	1.004	0.057	0.005	0.076	0.071
$N = 5000$; 1s = 16%	1.000	0.036	0.002	0.045	0.044
$N = 5000$; 1s = 36%	1.001	0.027	0.001	0.034	0.034
$N = 10{,}000$; 1s = 16%	0.998	0.025	0.001	0.032	0.032
$N = 10{,}000$; 1s = 36%	1.000	0.019	0.001	0.024	0.025
Cubic polynomials					
$N = 1000$; 1s = 16%	0.999	0.078	0.010	0.101	0.099
$N = 1000$; 1s = 37%	1.006	0.057	0.005	0.076	0.072
$N = 5000$; 1s = 16%	0.995	0.036	0.002	0.045	0.045
$N = 5000$; 1s = 36%	1.000	0.027	0.001	0.034	0.034
$N = 10{,}000$; 1s = 16%	0.992	0.026	0.001	0.032	0.032
$N = 10{,}000$; 1s = 36%	0.999	0.020	0.001	0.024	0.025
B-Splines					
$N = 1000$; 1s = 16%	1.005	0.078	0.010	0.101	0.100
$N = 1000$; 1s = 37%	1.007	0.057	0.005	0.076	0.072
$N = 5000$; 1s = 16%	1.000	0.036	0.002	0.045	0.045
$N = 5000$; 1s = 36%	1.001	0.027	0.001	0.034	0.034
$N = 10{,}000$; 1s = 16%	0.999	0.025	0.001	0.032	0.032
$N = 10{,}000$; 1s = 36%	1.000	0.020	0.001	0.024	0.025
Automated smoothing splines					
$N = 1000$; 1s = 16%	0.997	0.078	0.010	0.101	0.099
$N = 1000$; 1s = 37%	1.003	0.057	0.005	0.075	0.071
$N = 5000$; 1s = 16%	0.997	0.036	0.002	0.045	0.045
$N = 5000$; 1s = 36%	1.000	0.027	0.001	0.034	0.034
$N = 10{,}000$; 1s = 16%	0.995	0.026	0.001	0.032	0.032
$N = 10{,}000$; 1s = 36%	0.999	0.020	0.001	0.024	0.025

Note: Bias is the mean of absolute bias: $|\hat{\beta}_x - 1|$. MSE is the mean of expected squared bias: $E[(\hat{\beta}_x - 1)^2]$. SE is the mean of the simulated standard errors. SD is the standard deviation of the estimates.

link function described in equation (1).[17] I create four functional forms of hazard rates based on their relationship with time: increasing, decreasing, and two scenarios of non-monotonic relationships (one parabolic, one log-logistic).

We can assess the performance of the LTEs under different functional forms of the hazard rates with two simple criteria that directly inform the calculation of LTE: bias in the coefficient of the variable of interest (β_{X_K}), and by comparing the estimated hazards to the true hazard rate at different values of t. In the first three columns of Table 1, I show the average estimate, absolute bias, and mean squared error of $\beta_{X_K} = 1$ for five estimation techniques: exponential (flat, or no hazard), time dummies (Beck, Katz, and Tucker 1998), cubic polynomials (Carter and Signorino 2010), B-splines located at three knots (1, 4, and 7 to compare the default from many statistical programs), and automated smoothing splines (via generalized cross-validation). In the final two

[17]I follow the lead of Carter and Signorino (2010, 283) in designing these experiments.

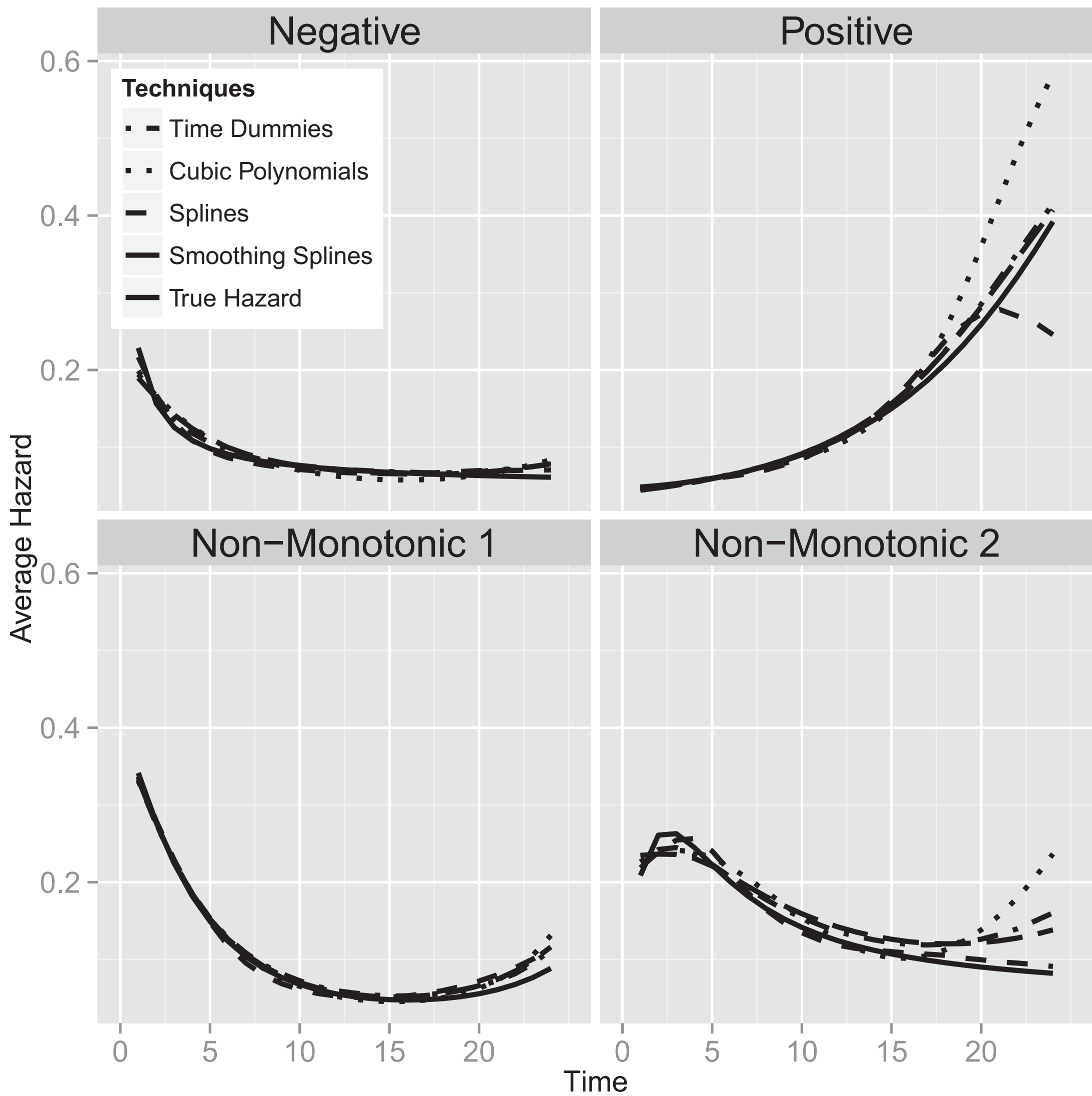

Fig. 3 Average hazard rates compared to true hazard rates.
Note: $N = 1000$, sims $= 1000$, beta $= c(-3, 1)$.

columns, I assess whether the standard errors accurately portray the variability of the estimate by comparing the mean standard error of β_{X_K} to the standard deviation of the 1000 estimates of β_{X_K}. Better estimators are those that have lower values of bias and where the average standard errors are closer to the standard deviations of the estimated β_{X_K} (Carsey and Hardin 2014, 84–96).

The evidence is clear; failing to correctly model temporal dependence biases the coefficient for the explanatory variable of interest.[18] In nearly every single scenario, the exponential distribution has the highest absolute bias and mean squared error. While the coefficients are biased downward, the similarity of the average standard error and the standard deviation of the coefficients means that the standard errors—in these circumstances[19]—reflect the true variability of the estimates. The

[18]This is consistent with omitted-variable bias present in probit models that ignore temporal dependence (Yatchew and Griliches 1985). In the Online Appendix, I show that this is the case for other functional forms as well (positive, parabolic, and log-logistic).

[19]Though this might be a function of the particular level of temporal dependence (which does not vary) in these experiments; other scholars find that the standard errors are often much smaller than they ought to be (Beck, Katz, and Tucker 1998, 1263).

evidence also suggests that as long as one specifies the temporal dependence in a reasonable fashion (thus excluding the exponential functional form), one is likely to retrieve the actual estimate of β_{X_K}, on average (Beck, Katz, and Tucker 1998, 1278–79; see also Carter and Signorino 2010, 284).

Of course, this is not to say that the LTE will be accurate under all circumstances, as this depends on appropriately modeling the underlying hazard rate as well. By comparing the true hazard rate to the estimates from the various techniques, we can assess whether the estimates contain large amounts of bias. Figure 3 provides these comparisons for four different functional forms estimated on a data set with 1000 observations and a constant of −3 (which produces a distribution where events occur in about 16% of the observations). The average hazard rates are quite close to the true hazard rate over the entire period, and the estimation techniques do well at picking up non-linearities in the hazard when they occur. The estimates reflect the slight increase at higher values in the first non-monotonic scenario and the curvilinear shape at the lower values of the second non-monotonic scenario.

More generally, the estimates are closer to the true hazard rate at lower, more common values of t.[20] As t increases, the estimates stray further from the true hazard, and this is exacerbated in two situations. First, when the constant increases (shown in the Online Appendix), there is a higher percentage of 1s in the sample (from about 15% to about 40%), at which point there is a greater divergence from the true hazard at lower values. Second, some techniques provide better estimates for some functional forms than others. For example, cubic polynomials do a particularly poor job at higher values of t, and they often depict non-linearities in the hazard rate that are not there (as shown in the *Non-Monotonic 2* scenario in Fig. 3).[21]

In practice, however, there is still reason to be optimistic about the use of these techniques to address temporal dependence. First, BTSCS models with temporal dependence are often predicting rather rare events such as civil wars or interstate conflicts, and these data sets are similar in attributes to the scenario depicted in Fig. 3. In situations where there is a lower percentage of 1s in the sample, scholars should have more confidence that a reasonable technique will provide a close estimate of the true hazard rate. Second, in the simulated data sets, values of t that are larger than about 20 are relatively rare (all of the data are right-skewed), so it is reasonable to expect that our estimates of the hazard rate under these anomalous situations would be potentially wrong. If scholars are cautious about making inferences about observations that are not representative of the entire sample, then they will focus their inferences on low to moderate values of t. In this range of values, the estimates are quite close to the truth.

In Fig. 2, I depicted how ignoring and modeling non-proportional hazards produces meaningfully different depictions of the long-run dynamics of a relationship. To identify the source of these differences, I estimate a series of Monte Carlo experiments on a similar data-generating process as described above.[22] The primary difference is that the influence of X_K now varies as a function of t ($X_K \times t$), and the influence varies across scenarios: ($\beta_{Xt} \in \{0.2, 0.1, 0.04, 0.02\}$). In the case of negative non-proportional hazards, these coefficients are negative.

In the Online Appendix, I interpret these experiments in greater detail, but a couple of patterns are informative. First, ignoring the non-proportional hazards biases the coefficient for X_K, and this bias increases with the degree of non-proportional hazards. Second, the estimated hazard rate tends to track the true hazard rate pretty closely, especially at lower values of t. Third, the estimates of the effects of X_K across values of t, or the average predictive differences (APD), demonstrate that failing to model minor levels of non-proportional hazards is not as problematic as one might think. In most cases, the estimated APD is able to capture the rising (declining) influence of a variable due to positive (negative) non-proportional hazards.

[20]This is also evident in the Monte Carlo experiments produced in the Online Appendix accompanying Carter and Signorino (2010, 7).

[21]These poor estimates occur even when one accounts for possible numerical instability problem by dividing the values of t by 100 before squaring and cubing them (Carter and Signorino 2010, 283).

[22]There are 1000 simulations of the estimates derived from a model with 1000 observations and coefficient values of −3 and 1 for the constant and β_{X_K}, respectively.

One interpretation of these findings is that there is little harm in estimating additive models that ignore non-proportional hazards. Of course, these results are somewhat limited to the unique circumstances present in this batch of experiments (with monotonic hazards with slight non-proportionality). If one believes that there are non-proportional hazards present, and if one's concern is on estimating the effects of the variable with non-proportional hazards, it still makes more sense theoretically to estimate the interactive model. Given the relative simplicity of estimating and testing for non-proportional hazards (particularly in the case of cubic polynomials), it is extremely short-sighted to move forward without a full examination of one's data. Indeed, this is the only way to ensure that the non-proportional nature of the relationship is properly understood.

Illustration

This section reveals how temporal dependence influences the inferences that one makes from their BTSCS model. Each simulation scenario used to depict quantities of interest has a probabilistic LTE made up of two components: the change in probability of the outcome occurring at time t and the change in the probability of the outcome in future time periods, given a conflict at time t. Since influencing the risk of the outcome at time t modifies future values of the temporal dependence variables, scholars underestimate the substantive effects by neglecting the LTEs. In this example, Clare's (2010) key theoretical variable has a long-lasting effect that extends much farther than we currently believe. There is reason to believe that the total effects have been seriously understated in Clare (2010).

Clare (2010) theorizes that the behaviors of parliamentary governments vary based on the extent to which the governing coalition is cohesive. The presence of outlier parties—either on the left or the right—will have a disproportionate influence on foreign policy behavior because they can threaten to leave the coalition. Since Clare (2010, 979) includes *peaceyears*, *peaceyears*^{2} and *peaceyears*^{3} to "address the problem of temporal dependence," this project offers a perfect opportunity to explore the effects of including temporal dependence variables.[23]

Since Clare's (2010) focus is on how ideological outliers pressure coalition governments into either peaceful or adventurous foreign policies, I begin by calculating the LTEs of an increase in *ideological distance* from its minimum value of -110 (representing a far-left outlier coalition partner) to $+122$ (representing a far-right outlier coalition partner). The first step is to establish the simulation scenario ($\mathbf{X_C}$). Each simulation scenario ($\mathbf{X_C}$) used to depict quantities of interest provides both a baseline probability of the event occurring at time t ($\Pr(y_i = 1|\mathbf{X_C})$) and the information needed to calculate the LTE (see equations (3)–(4)). In this case, I treat each observation in the data set as a separate simulation scenario so that I can quantify the effects of *ideological distance* given observed values of the independent variables (Hanmer and Kalkan 2013).

Figure 4a shows the probability of dispute initiation given the observed simulation scenario combined with the minimum value of *ideological distance* ($\Pr(\hat{y} = 1|\mathbf{X_C}, \mathrm{ID} = \min)$), across the range of values of *peaceyears*. Recall that the predicted probability provides two pieces of vital information: first, the probability that the event occurs at time t, and second, the probability of a LTE at future values of t. While the first is widely recognized and displayed as the principal quantity of interest, the second is crucial for a complete depiction of the substantive effects. The probabilities are quite small (ranging from nearly 0 to 0.05), which is common in models of rare events (King and Zeng 2001).[24]

Figure 4b provides the probability of the event occurring at time t, given an increase in *ideological distance* to its maximum ($\Pr(\hat{y} = 1|\mathbf{X_C}, \mathrm{ID} = \max)$). The difference between the values in the two panels represents the probabilistic component of the LTE ($\Delta\Pr(\hat{y} = 1|\mathbf{X_C}, \Delta\mathrm{ID})$. The first inference is that increasing *ideological distance* has by far its biggest impact on the probability of dispute initiation (and, importantly, the probability of an LTE) if a conflict recently occurred. Keep

[23]All of the following discussion is based on Model 1 in Table 1 (Clare 2010, 982).

[24]The model fits the data reasonably well, as the ePCP is 94.6% and the area under the curve (AUC) is 0.77 (with a 95% confidence interval of [0.71, 0.82]). In the Online Appendix file, I also provide the ROC and separation plots.

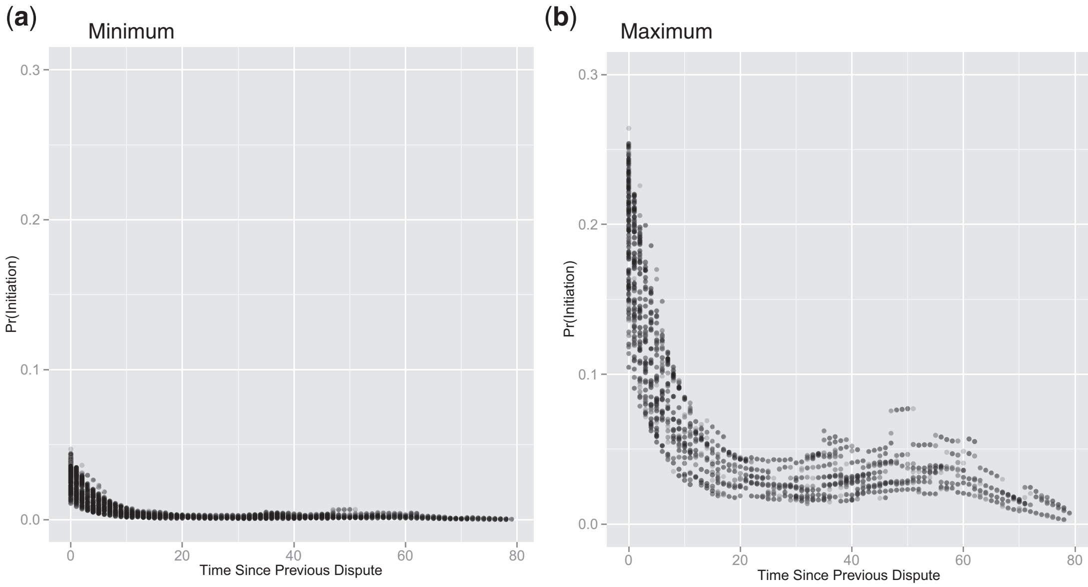

Fig. 4 Probabilities of LTEs when *ideological distance* is at its minimum and its maximum values. *Note:* Scatterplots are slightly jittered to illustrate the density of observations. Each dot reflects the probability of an LTE given the actual values for that observation.

in mind that, since *ideological distance* and *peaceyears* are incorporated additively, these varying effects are due solely to compression.[25]

The next step in depicting LTEs is to calculate the change in probability of the outcome at time $t+1$ to $t+k$, given an outcome occurring at time t (see equations (3)–(4)). This can be done in a variety of ways, though I present two here. The first option is to calculate the "average" LTEs with the observed-value approach (Hanmer and Kalkan 2013) using the following steps:

1. Using the observed values for observation #1, calculate the change in probability of a dispute initiation given a change in an explanatory variable.
2. Now assume that a dispute occurs at time t; calculate the change in probability at time $t+1$ when *peaceyears* (and its squared and cubed terms) is reset to 0 compared to what it would be for that observation without a dispute at time t. Repeat this step for values $t+1$ through $t+k$.
3. Repeat steps #1–2 for all N observations in the data set.
4. At each value of $t+1$ through $t+k$, calculate the mean of those N LTEs to generate the "average" LTEs with the observed-value approach.

Figure 5 shows the probabilities of conflict for the two counterfactual scenarios (Fig. 5a) and the "average" LTE (Fig. 5b) from $t+1$ through $t+20$. Experiencing a conflict at time t immediately increases the risk of a dispute, which is consistent with a wide range of theories of international conflict, for example, the lasting effects of territorial claims on the potential for future hostilities (e.g., Senese 2005). Figure 5b shows that the LTE of conflict at time t produces a statistically significant increase for the next 15 time periods. Furthermore, while the absolute values of the LTE appear small (0.09 at $t+1$), the LTEs are quite large relative to the baseline probability at time t (0.11).

An alternative to depicting the "average" LTEs is to establish multiple simulation scenarios where the baseline probability varies considerably. This shows the influence of hyper-conditionality

[25]In the Online Appendix file, I demonstrate that we fail to reject the null of proportional hazards, which means that any change in the substantive effects of *ideological distance* across *peaceyears* is due to compression rather than an omitted interactive relationship (see Carter and Signorino 2010).

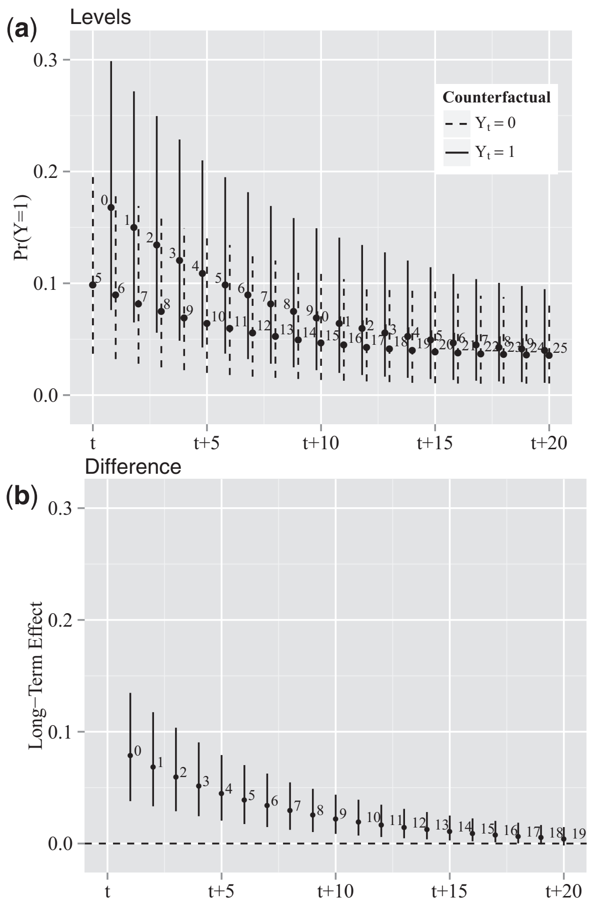

Fig. 5 Average LTEs across *time since previous dispute*.
Note: Lines depict 95% confidence intervals for the LTEs calculated using the observed-values approach. Numbers attached to points represent the values of *t* in the scenario.

by allowing the size of the substantive effect to vary according to its position along the CDF. Instead of crafting simulation scenarios that reflect substantively interesting cases, for the purposes of this illustration, I create scenarios where the baseline probability of dispute initiation varies considerably: 0.003, 0.05, 0.10, and 0.50.[26] The value of *peaceyears* at time *t* is 5 across all four scenarios. Figure 6 plots the LTEs of a dispute at time *t* (with 95% confidence intervals) for the first eleven years following the dispute for these four scenarios.

As expected due to compression, the substantive and statistical signficance of the LTEs varies according to the baseline probabilities. As the baseline probability increases (across panels of Fig. 6), the magnitude of the LTE increases considerably. Moreover, the LTE is statistically different from 0 at all the depicted values of *peaceyears*. Given the massive influence that hyperconditionality can have on the magnitude and statistical significance of one's inferences, these figures would point to the utility of evaluating these LTEs at multiple simulation scenarios of varying baseline probabilities. Otherwise, one might be vulnerable to offering general inferences that only hold for a small portion of the observations.

[26]It should be noted that the fourth scenario is highly unrealistic given the estimated parameters (the maximum in-sample probability is 0.16) and is only intended to demonstrate how the LTE are influenced by compression. In the Online Appendix, I provide the values of the variables selected for the scenarios.

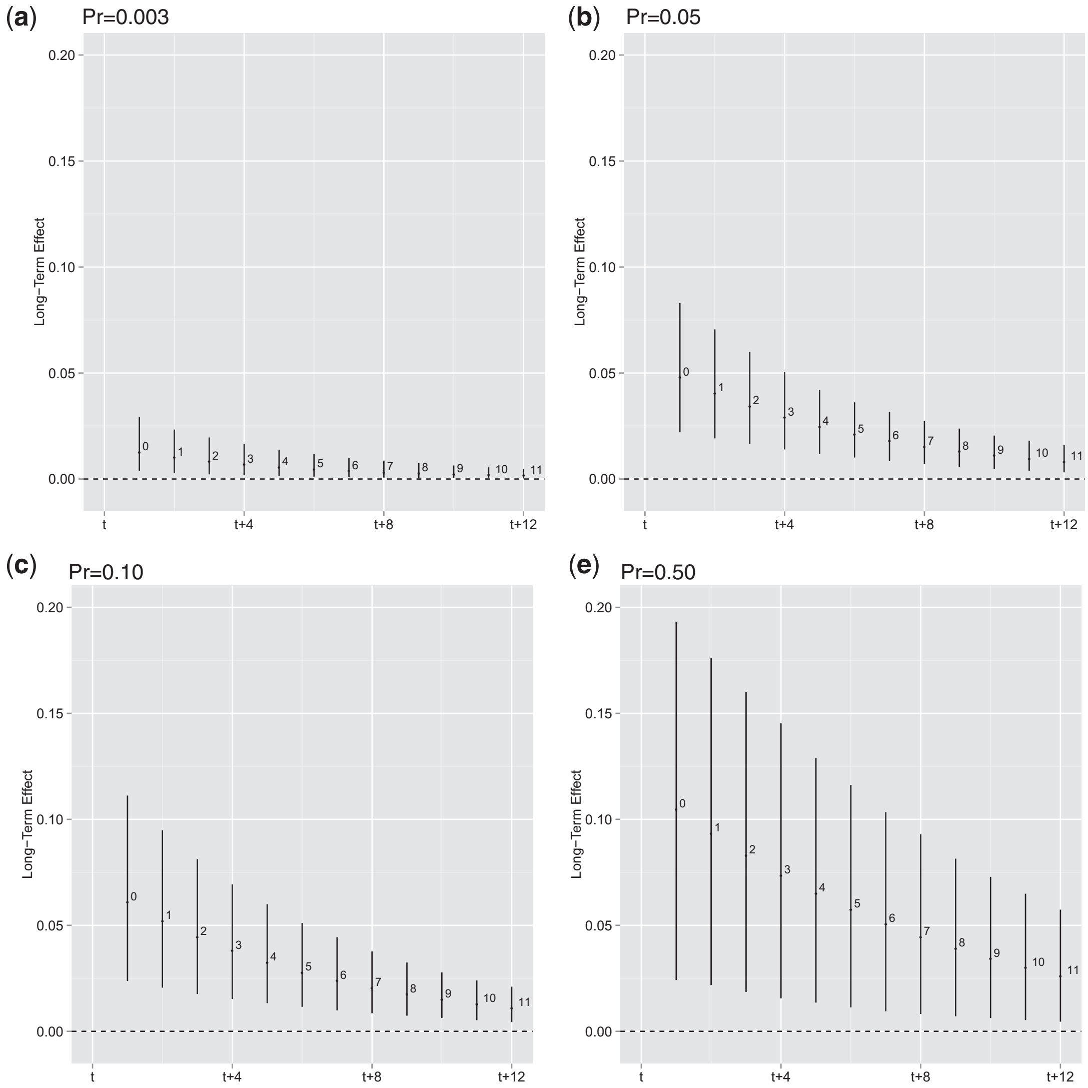

Fig. 6 LTEs across *time since previous dispute* for four simulation scenarios with varying baseline probabilities.

Note: Lines depict 95% confidence intervals for the LTEs. Probabilities denote the baseline probability of a dispute initiation for that simulation scenario. Numbers attached to points represent the values of t in the scenarios.

Conclusion

The goal of this project is to convince scholars that including temporal dependence variables does much more than just provide peace of mind to those worried about temporal dependence. Rather, including these variables—which often have massive influences on the outcome of interest—changes the substantive effects of the key theoretical variables. Indeed, the inclusion of time counters, cubic splines, or polynomials allows variables to have an LTE in addition to the immediate effect. Prominent theories (such as the conflict trap; see Collier et al. 2003) often have expectations that the variables have long-lasting effects, or that the a variable's influence grows with each recurring event. Yet, up until this point, scholars have been unable to estimate and graphically depict these theoretically interesting long-run dynamics from BTSCS models. I presented a wide range of quantities of interest one can derive from BTSCS models with temporal dependence, ranging from responses to a variety of changes in X, compounded effects, and effects that vary with time. Putting these principles into practice is rather easy. Correctly interpreting the LTEs requires no change in the estimation technique, and only requires adjusting how one interprets the substantive effects. Interpreting one's model in this manner allows scholars to paint a more complete picture of the causal story under investigation.

I demonstrated the importance of these effects by replicating Clare's (2010) study of conflict behavior of parliamentary democracies. I showed that variables can be much more influential than previously thought because of the potential to change future values of the temporal dependence variables. The inferences one makes about the LTEs of *ideological distance* vary in magnitude and statistical significance based on the values chosen for the temporal dependence variables. I then offered two approaches to illustrating the LTE across multiple scenarios so that scholars can observe the consequences of those decisions. In a companion piece, Gandrud and Williams (2016) introduce R software that automates the process of generating probabilistic LTEs in binary models with temporal dependence.

While this project demonstrated these effects with an example of a probit model with cubic polynomials, the intuition is broadly applicable to different ways of operationalizing temporal dependence, such as various kinds of splines. Since these methods depend on the time since the occurrence of the event, it is simple to change the future values of the spline for the applicable observations. Furthermore, this intuition extends to other models that employ temporal dependence variables, such as count models (Li 2009), ordered logits (Melander 2005), multinomial logits (Buhaug and Gleditsch 2008), and conditional logits (Dreher and Gassebner 2012). In all of these models, explanatory variables potentially have a long-term component. The difference arises in how one determines whether an "event" occurs, through either probabilities (in the case of conditional logit, ordered logit, or multinomial logit) or expected values (in the case of count models).

Conflict of interest statement. None declared.

References

Ai, Chunrong, and Edward C. Norton. 2003. Interaction terms in logit and probit models. *Economics Letters* 80:123–29.

Beardsley, Kyle. 2008. Agreement without peace? International mediation and time inconsistency problems. *American Journal of Political Science* 52:723–40.

Beck, Nathaniel. 2001. Time-series–cross-section data: What have we learned in the past few years? *Annual Review of Political Science* 4:271–93.

———. 2010. Time is not a theoretical variable. *Political Analysis* 18:293–94.

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* 42:1260–88.

Bennett, D. Scott. 1999. Parametric models, duration dependence, and time-varying data revisited. *American Journal of Political Science* 43:256–70.

———. 2006. Toward a continuous specification of the democracy-autocracy connection. *International Studies Quarterly* 50:313–38.

Bennett, D. Scott, and Allan Stam. 1996. The duration of interstate wars, 1816–1985. *American Political Science Review* 90:239–57.

———. 2000. Research design and estimator choices in the analysis of interstate dyads: When decisions matter. *Journal of Conflict Resolution* 44:653–85.

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* 54:248–66.

Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event history modeling: A guide for social scientists*. Cambridge: Cambridge University Press.

Buhaug, Halvard, and Kristian Skrede Gleditsch. 2008. Contagion or confusion? Why conflicts cluster in space. *International Studies Quarterly* 52:215–33.

Carter, David B., and Curtis S. Signorino. 2010. Back to the future: Modeling time dependence in binary data. *Political Analysis* 18:271–92.

Clare, Joe. 2010. Ideological fractionalization and the international conflict behavior of parliamentary democracies. *International Studies Quarterly* 54:965–87.

Collier, Paul, V.L. Elliot, Havard Hegre, Anke Hoeffler, Marta Reynal-Querol, and Nicholas Sambanis. 2003. *Breaking the conflict trap*. Washington, D.C.: Oxford University Press.

Crescenzi, Mark J. C. 2007. Reputation and international conflict. *American Journal of Political Science* 51:382–96.

de Boef, Suzanna, and Luke Keele. 2008. Taking time seriously: Dynamic regression. *American Journal of Political Science* 52:184–200.

Desposato, Scott W. 2006. Parties for rent? Ambition, ideology, and party switching in Brazil's chamber of deputies. *American Journal of Political Science* 50:62–80.
Esarey, Justin, and Jacqueline H. R. DeMeritt. 2014. Defining and modeling state-dependent dynamic systems. *Political Analysis* 22:61–85.
Gandrud, Christopher, and Laron K. Williams. 2016. *pltesim: An R package for simulating probabilistic long-term effects in models with temporal dependence*. Working Paper, Hertie School of Governance.
Greene, William H. 2008. *Econometric analysis*. 6th ed. Upper Saddle River, NJ: Pearson Prentice Hall.
Greenhill, Brian, Michael D. Ward, and Audrey Sacks. 2011. The separation plot: A new visual method for evaluating the fit of binary models. *American Journal of Political Science* 55:991–1003.
Grieco, Joseph M. 2001. Repetitive military challenges and recurrent international conflicts, 1918–1994. *International Studies Quarterly* 45:295–316.
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* 57:263–77.
Herron, Michael C. 1999. Post-estimation uncertainty in limited dependent variable models. *Political Analysis* 8:83–98.
Jackman, Simon. 2000. *In and out of war and peace: Transitional models of international conflict*. Working Paper, Stanford University.
Keele, Luke. 2008. *Semiparametric regression for the social sciences*. West Sussex, England: Wiley and Sons.
King, Gary, and Langche Zeng. 2006. The dangers of extreme counterfactuals. *Political Anaysis* 14:131–59.
King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science* 44:347–61.
Li, Quan. 2009. Democracy, autocracy, and expropriation of foreign direct investment. *Comparative Political Studies* 42:1098–1127.
Licht, Amanda A. 2011. Change comes with time: substantive interpretation of nonproportional hazards in event history analysis. *Political Analysis* 19:227–43.
Londregan, John B., and Keith T. Poole. 1990. Poverty, the coup trap, and the seizure of executive power. *World Politics* 42:151–83.
Long, J. Scott. 1997. *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage Publications.
McGrath, Liam F. 2015. Estimating oonset of binary events in panel data. *Political Analysis* 23:534–49.
Melander, Erik. 2005. Gender equality and intrastate armed conflict. *International Studies Quarterly* 49:695–714.
Mooney, Christopher Z., and Mei-Hsien Lee. 2000. The influence of values on consensus and contentious morality policy: U.S. death penalty reform, 1956–82. *Journal of Politics* 62:223–39.
Nagler, Jonathan. 1991. The effect of registration laws and education on United States voter turnout. *American Political Science Review* 85:1393–1405.
———. 1994. Scobit: An alternative estimator to logit and probit. *American Journal of Political Science* 38:230–55.
Norton, Edward C., Hua Wang, and Chunrong Ai. 2004. Computing interaction effects and standard errors in logit and probit models. *The Stata Journal* 4:154–67.
Oneal, John R., and Bruce Russett. 1999. Assessing the liberal peace with alternative specifications: Trade still reduces conflict. *Journal of Peace Research* 36:423–42.
Pindyck, Robert S., and Daniel L. Rubinfeld. 1991. *Econometric models and economic forecasts*. 3rd ed. New York: McGraw-Hill.
Przeworski, Adam, and James Raymond Vreeland. 2002. A statistical model of bilateral cooperation. *Political Analysis* 10:101–12.
Przeworski, Adam, Michael E. Alvarez, Jose Antonio Cheibub, and Fernando Limongi. 2000. *Democracy and development: Political institutions and well-being in the world, 1950–1990*. Cambridge: Cambridge University Press.
Rainey, Carlisle. 2015. *Transformation-induced bias: Unbiased coefficients do not imply unbiased quantities of interest*. Working Paper, Texas A&M University.
Senese, Paul D. 2005. Territory, contiguity, and international conflict: Assessing a new joint explanation. *American Journal of Political Science* 49:769–79.
Shipan, Charles R., and Craig Volden. 2008. The mechanisms of policy diffusion. *American Journal of Political Science* 52:840–57.
Simmons, Beth A. 2000. International law and state behavior: Commitment and compliance in international monetary affairs. *American Political Science Review* 94:819–35.
Smith, Benjamin. 2004. Oil wealth and regime survival in the developing world, 1960–1999. *American Journal of Political Science* 48:232–46.
Thompson, William R. 2001. Identifying rivals and rivalries in world politics. *International Studies Quarterly* 45:557–86.
Train, Kenneth E. 2009. *Discrete choice methods with simulation*. 2nd ed. Cambridge, England: Cambridge University Press.
Weidmann, Nils B., and Michael D. Ward. 2010. Predicting conflict in space and time. *Journal of Conflict Resolution* 54:883–901.
Williams, Laron K. 2016. Replication Data for: Long-term effects in models with temporal dependence. http://dx.doi.org/10.7910/DVN/KRKWK8,HarvardDataverse,V1 [UNF:6:pwv9cSHI53tZqXlrJ9EDaw¼¼.
Williams, Laron K., and Guy D. Whitten. 2012. But wait, there's more! Maximizing substantive inferences from tscs models. *Journal of Politics* 74:685–93.
Zorn, Christopher. 2000. Modeling duration dependence. *Political Analysis* 8:367–80.