

# Back to the Future: Modeling Time Dependence in Binary Data

David B. Carter  
Curtis S. Signorino\*

Department of Political Science  
University of Rochester

March 2, 2009

## Abstract

Since Beck, Katz, and Tucker (1998), the standard method for modeling time dependence in binary data has been to incorporate time dummies or splined time in logistic regressions. Although we agree with the need for modeling time dependence, we demonstrate that time dummies can induce serious estimation problems due to separation. Splines do not suffer from these problems. However, the complexity of splines has led substantive researchers (1) to use knot values that may be inappropriate for their data and (2) to ignore any substantive discussion concerning the effect of time. We propose a relatively simple alternative: including  $t$ ,  $t^2$ , and  $t^3$  in the regression. This cubic polynomial approximation is trivial to implement — and, therefore, interpret — and it avoids problems such as quasi-complete separation. Monte Carlo analysis demonstrates that, for the types of hazards one often sees in substantive research, the polynomial approximation always outperforms time dummies and generally performs as well as splines, or even more flexible auto-smoothing procedures. Due to its simplicity, this method also accommodates nonproportional hazards in a straightforward way. We reanalyze Crowley and Skocpol (2001) using nonproportional hazards and find new empirical support for their theory.

---

\*We thank Hein Goemans, Luke Keele, Eduardo Leoni, Arthur Spirling, Randy Stone, and Chris Zorn for their comments and suggestions, as well as the participants in the Watson Center for Conflict and Cooperation Seminar Series. We also thank Jocelyn Crowley, Theda Skocpol, Matt Dickinson, and Glenn Palmer for providing their data. We gratefully acknowledge support from the National Science Foundation (SES-0213771, SES-0413381). Emails: [dcarter5@mail.rochester.edu](mailto:dcarter5@mail.rochester.edu), [curt.signorino@rochester.edu](mailto:curt.signorino@rochester.edu).

# 1 Introduction

Whether it is the primary focus or not, political scientists are often interested in how the phenomena they study change over time. For example, are democracies more stable the longer they have been established? Are congressional incumbents more likely to survive an election the longer they have been in office? Are two recently warring nations more or less likely to become embroiled in another conflict as time goes on? In answering questions such as these, we should first think theoretically about the mechanisms that cause our subject of interest to change (or not) over time. Furthermore, in our empirical analysis we need a method that is flexible enough to allow for a variety of theoretical relationships between time and the phenomenon being studied.

Increasingly, researchers have access to refined (or higher resolution) versions of event history data. In one common form of this data, a binary dependent variable represents whether an event (such as government transition or war occurrence) occurred or not during some slice of time. First advocated by Beck, Katz and Tucker (1998), logistic regression with time dummies or splined time has become the standard method for analyzing this type of data. Researchers who study a wide variety of topics in international relations, American politics, and comparative politics have all adopted the Beck, Katz, and Tucker (1998) (hereafter, BKT) recommendations.

We should be clear in stating from the outset that we completely agree with BKT's main point: scholars should "take time seriously".<sup>1</sup> However, the vast majority of researchers have treated temporal dependence in binary data models more as a statistical nuisance that needs to be "controlled for," rather than as something that is substantively interesting. Indeed, most of those who have followed BKT's advice subsequently ignore time in discussions of empirical results.

Consider Table 1, which lists all published articles we found that follow BKT's advice on using time dummies or splines.<sup>2</sup> We have further classified the citations according to whether the authors interpreted the effect of time or not. Table 1 demonstrates that despite both splines and time dummies being extensively used in every substantive field of political science, virtually no one actually plots and interprets the hazard. In fact, out

---

<sup>1</sup>See, for example, Alt, King and Signorino (2001).

<sup>2</sup>We compiled this list by locating all published articles that cited BKT in the Social Sciences Citation Index (SSCI) as of July, 2006. We then went through all the articles to determine whether they implemented either splines or time dummies and whether they interpreted the hazard.

of 91 studies that utilize splines, only 3 actually plot and interpret a hazard. The track record for time dummies is slightly better, but out of 28 studies that utilize dummies, only 4 plot the hazard. In short, the discipline’s track record shows that rarely is time taken seriously in a substantial way.

What accounts for researchers taking BKT seriously, but not time? We suspect it stems from the difficulties in either the implementation or interpretation of the two methods they propose. As we will show, serious practical problems can arise in the implementation of time dummies. Researchers can avoid those problems by using splines. Moreover, researchers do not even need to know how to generate the splined time data. To assist substantive researchers, BKT provided a Stata routine to generate splined time variables to include in regression. However, most political scientists do not seem to understand what splines are.<sup>6</sup> The default parameters (e.g., “knot” values) are often used, regardless of whether they are appropriate for the researcher’s data. And, as Table 1 shows, the relationship between the dependent variable and time is rarely ever discussed when splines are employed.

In this paper, we propose a simpler alternative that has advantages in terms of both modeling and interpreting time dependence: using  $t$ ,  $t^2$ , and  $t^3$  in one’s regression, which serves as a third-order Taylor series approximation to the hazard. In contrast to time dummies, this cubic polynomial approximation does not cause the same data problems.<sup>7</sup> The cubic polynomial is related to splines, but much easier to implement and interpret. Indeed, modeling and interpreting nonproportional hazards is relatively straightforward with the cubic polynomial approach.

This paper proceeds as follows. In the next section, we briefly discuss the link between duration models and their binary data equivalents. Following that, we examine the implementation and interpretation issues with time dummies and splines. In Section 5, we show via Monte Carlo analysis that our method outperforms time dummies and generally performs as well as splines. Following that, we replicate the original findings in Crowley and Skocpol (2001) and then extend their analysis using a nonproportional hazards model.

---

<sup>3</sup>The authors do not show the hazard in the text of the article but do offer an appendix with a plot of the hazard upon request.

<sup>4</sup>They do not actually plot the full hazard but do report the hazard for specific years in a table.

<sup>5</sup>The authors include a table that shows the effect of two covariates on the hazard, but do not show the hazard as a function of time.

<sup>6</sup>See, however, Keele (2008).

<sup>7</sup>We use the terms “cubic polynomial” and “ $t$ ,  $t^2$ , and  $t^3$ ” interchangeably throughout the paper.

Table 1: Use of Splines and Time Dummies

	No Hazard	Interpret Hazard	
<b>Splines</b>	<b>88 (96.7%)</b>	<b>3 (3.3%)</b>	
	<p>Goodcliffe and Hawkins (2006)<sup>3</sup>  Hafner-Burton and Montgomery (2006)  Dorussen (2006)  Caprioli and Trumbore (2006)  Rasler and Thompson (2006)  Mansfield and Pevehouse (2006)  Braithwaite (2005)  Buhaug (2005)  Benson (2005)  Oneal and Russett (2005)  Meinke, Staton and Wuhs (2006)  Bearce and Omori (2005)  Edwards (2005)  Melander (2005)  Enterline and Greig (2005)  Barbieri and Reuveny (2005)  Chamberlain and Haider-Markel (2005)  Kim and Rousseau (2005)  Caprioli and Trumbore (2005)  Senese (2005)  Besancon (2005)  Humphries (2005)  Lujala, Gleditsch and Gilmore (2005)  Marinov (2005)  Caprioli (2005)  Sobek (2005)  Milner and Kubota (2005)  Boehmer and Sobek (2005)  Sørli, Gleditsch and Strand (2005)  Fritz and Sweeney (2004)  Krause (2004)  Powers (2004)  Mitchell and Prins (2004)  Gowa and Mansfield (2005)  Sechser (2004)  Rasler and Thompson (2004)  Goenner (2004)  McDonald (2004)  Volden and Carruba (2004)  Walter (2004)  Schamis and Way (2003)  Sweeney and Fritz (2003)  Cauthen and Peters (2003)  Lai (2003a)</p>	<p>Leblang and Chan (2003)  Ray (2003)  Reed (2003b)  Oneal (2003)  Gartzke and Li (2003c)  Gartzke and Li (2003a)  Sweeney (2003)  Choi and James (2003)  Mansfield and Reinhardt (2003)  Gartzke and Li (2003b)  Ravlo, Gleditsch and Dorussen (2003)  Senese and Vasquez (2003)  Bennett and Rupert (2003)  Montinola (2003)  Lai (2003b)  Reed (2003a)  Clark and Regan (2003)  Chiozza (2002)  Pickering (2002)  Kinsella and Russett (2002)  Jungblut and Stoll (2002)  Davies (2002)  Dixon and Senese (2002)  Mousseau (2002)  Colaresi and Thompson (2002)  Peceny, Beer and Sanchez-Terry (2002)  Li and Sacko (2002)  Zorn (2001)  Rasler and Thompson (2001)  Drury (2001)  Mousseau (2001)  Crescenzi and Enterline (2001)  Sherman (2001)  Henderson and Tucker (2001)  Gartzke, Li and Boehmer (2001)  Bennett and Stam (2000a)  Lektzian and Souva (2001)  Lemke and Reed (2001)  Peet and Simon (2000)  Mansfield and Pevehouse (2000)  Bennett and Stam (2000b)  Mousseau (2000)  Oneal and Russett (1999b)  Oneal and Russett (1999a)</p>	<p>Gelpi and Grieco (2001)  Simmons (2000)  Beck, King and Zeng (2000)</p>
<b>Time Dummies</b>	<b>24 (85.7%)</b>	<b>4 (14.3%)</b>	
	<p>Brinks and Coppedge (2006)  Volden (2006)  Stein (2005)  Chang (2005)  Krutz (2005)  Heath (2005)  Lebovic (2004)  Stalley (2003)  Arceneaux (2003)  Lebovic (2003)  Gelpi and Feaver (2002)  Howard and Roch (2001)  Henisz (2002)  Dickinson and Tempas (2002)  Ka and Teske (2002)  Volden (2002)  Ka and Teske (2002)  Crowley and Skocpol (2001)<sup>5</sup>  Balla (2001)  Mooney (2001)  Mooney and Lee (2000)  Palmer and Whitten (2000)  Reed (2000)  Thacker (1999)  Leblang (1999)</p>	<p>James (2006)  Carpenter and Lewis (2004)  Bernard, Reenock and Nordstrom (2003)<sup>4</sup>  Clark and Hart (1998)</p>	

Our nonproportional hazards version provides new, richer empirical support for Crowley and Skocpol (2001)'s theory.

## 2 Time Dummies for Dummies

The starting point for BKT is the important observation that, increasingly, the binary data we use in political science is a disaggregated (or less aggregated) form of event history data. Although BKT refer to this data as binary time-series cross-section (BTSCS) data, the focus is really on temporal dependence, rather than cross-sectional interdependence. We similarly focus on time here.

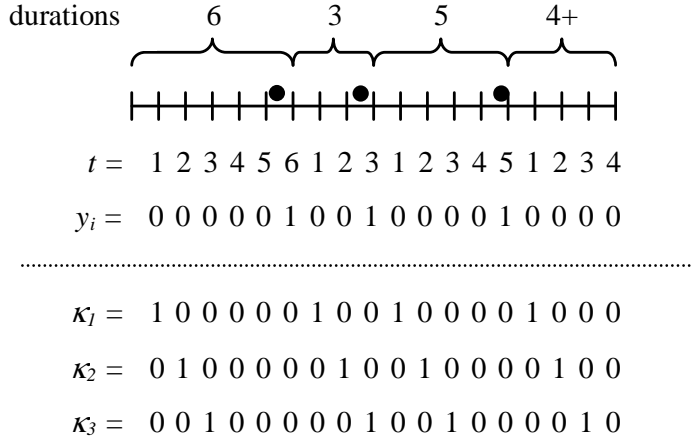
To help make this more concrete, consider the upper half of the time line of events displayed in Figure 1. Here, we have three events, denoted by the black dots. The durations, or time between successive events, are shown along the top: 6, 3, and 5. If we were analyzing duration data, our observations would generally correspond to these, along with the last right-censored observation of length 4. Vast literatures exist on duration (or survival) analysis (see, for example, Box-Steffensmeier and Jones (2004)). Rather than reviewing that literature, we simply point out that there are well-known techniques for modeling temporal dependence in duration data. Parametric models like the Weibull, log-logistic, or log-normal allow the analyst to estimate whether the hazard is increasing, decreasing, or non-monotonic with time. Alternatively, researchers sometimes opt for the (semi-parametric) Cox model.

Now consider the binary data shown below the timeline and denoted as  $y_i$ . As BKT have noted (and Alt, King and Signorino (2001) have demonstrated), if the data generating process is temporally dependent, then the use of a model such as logit with only a linear  $x\beta$  specification is inappropriate, since it implies a constant hazard. The question then becomes one of how to allow for temporal dependence in binary data without being too restrictive concerning the form of that dependence.

As BKT show, derivation of the binary data version of a Cox model is relatively straightforward (Beck, Katz and Tucker, 1998; Meyer, 1990; Narendranathan and Stewart, 1993; Prentice and Gloeckler, 1978; Katz and Sala, 1996). If one starts with the assumption of a Cox relative risk model with hazard

$$h(t|x_{i,t}) = h_0(t) \exp(x_{i,t}\beta) \tag{1}$$

Figure 1: Binary Representation of Duration Data



then the equivalent binary data model is a complementary log-log (cloglog) model, which can be written as

$$Pr(y_{i,t} = 1) = 1 - \exp(-\exp(x_{i,t}\beta + \kappa_{i,t}\alpha_t)) \tag{2}$$

where  $t$  indicates the time since the last event (or the life of the current subject),  $\kappa_{i,t}$  is a time dummy for period  $t$ , and  $\alpha_t$  is the effect of period  $t$  on the probability of an event occurring.<sup>8</sup>

It is important to note that the “time dummies” are not just the time counter  $t$ . Returning to Figure 1, we have displayed three time dummies below the dotted line. For example,  $\kappa_1$  will be one whenever  $t = 1$  and zero otherwise. Similarly,  $\kappa_2$  will equal one whenever  $t = 2$  and zero otherwise. In general, one needs a  $\kappa_t$  for every  $t$  in the data (although more about this later). If the time unit of analysis is the year and the longest duration in the data is 10 years, then there will be ten time dummies. If it is 30, then there will be 30 time dummies. The researcher can include all time dummies and drop the constant or include the constant and drop one time dummy. Based on this formulation, it is easy to see that if time dependence is present in data and it is not modeled, then

<sup>8</sup>It is a slight abuse of terminology to refer to the plots of  $Pr(y_{i,t} = 1)$  as plots of the “hazard.” Rather than an instantaneous failure rate, it should really be thought of as the approximate hazard over one-unit periods. As the aforementioned proofs demonstrate (e.g., in BKT’s appendix), the relationship between  $Pr(y_{i,t} = 1)$  and time is identical to that of the hazard, upon which it is based. An increasing hazard produces a  $Pr(y_{i,t} = 1)$  that increases with time; a decreasing hazard produces a  $Pr(y_{i,t} = 1)$  that decreases with time; and so on for nonmonotonic hazards.

omitted variable bias is present, which can bias other coefficients of interest in commonly used binary dependent variable models even when the omitted variables are not correlated with included variables (Yatchew and Griliches, 1985).

The cloglog model in equation 2 has the nice feature that we do not have to assume any particular hazard shape (e.g., increasing, decreasing, or non-monotonic). In principle, this allows consistent  $\beta$  estimates, even if the true shape of the hazard is unknown (Meyer, 1990). As BKT demonstrate, the logit model very closely approximates the cloglog. Moreover, there is no reason to privilege the cloglog distribution over logit. Therefore, BKT recommend that researchers use logit (or probit), since they are so widely available.<sup>9</sup> Like Cox models, time dummies can produce a “bumpy” hazard plot. BKT, therefore, advocate the use of natural cubic splines when researchers want a relatively smooth plot of the hazard.

## 2.1 Potential Problems with Time Dummies

The fact that the time dummies model can be derived directly from the Cox model would suggest that it should be the “go to” model for researchers. However, there are at least two potential problems with the time dummies: inefficiency and separation. The first problem, inefficiency, is clear from the preceding explication of the time dummies model: the greater the number of time periods in the data, the greater the number of parameters that will need to be estimated. 30 years in the data will be represented by 30 time dummy parameters. This is relatively inefficient compared to splines or the cubic polynomial approach — the latter requiring estimation of only three additional parameters. Of course, as with any estimation problem, the inefficiency is a matter of degree. It will be worse when the researcher needs to estimate many time dummies but has a small sample of data. It will likely be negligible when fewer time dummies are estimated or the sample is very large.

The issue of separation is potentially more problematic, especially when the dependent variable and regressors are dummy variables. In binary data, separation occurs when values (or ranges of values) of our regressors perfectly determine whether  $y_i$  is zero or one. In practice, individual dummy regressors are most often responsible for separation. However, separation need not be the result of a single variable. In general, data are completely separated if there exists some vector of coefficients  $\beta$ , such that for all observations  $\beta' x_i > 0$

---

<sup>9</sup>This is not really an issue anymore, as cloglog is widely available in statistical packages such as Stata, SPSS, and R. However, we will maintain the use of logit throughout this article.

Obs $i$	$t$	$\kappa_1$	$\kappa_2$	$\kappa_3$	$y_i$
1	1	1	0	0	1
2	1	1	0	0	0
3	2	0	1	0	1
4	1	1	0	0	0
5	2	0	1	0	0
6	3	0	0	1	1

Table 2: Separation in Binary Data with Three Time Dummies

when  $y_i = 1$  and  $\beta' x_i < 0$  when  $y_i = 0$ . Quasi-complete separation holds when these conditions are replaced by weak inequality, when at least one observation is satisfied by equality, and the  $\beta$  vector is not zero. If neither complete separation, nor quasi-complete separation hold, then the data are said to “overlap” (Albert and Anderson, 1984; Santner and Duffy, 1986).<sup>10</sup>

These three cases, complete separation, quasi-complete separation, and overlap, are mutually exclusive. However, only in the case of overlap do non-infinite maximum likelihood estimates (MLE) exist.<sup>11</sup> If the data exhibit either complete or quasi-complete separation, no maximum likelihood estimates exist and MLE routines will push the value of the estimated  $\hat{\beta}$  for the offending regressor(s) to  $\infty$  or  $-\infty$ . Many canned logit/probit procedures in commonly used software such as Stata will automatically check for separation in the data and drop any offending variables and observations stating that the variable “perfectly predicts”  $y$  in some number of cases. Alternatively, researchers may opt for more complicated estimation methods such as penalized maximum likelihood (Firth, 1993; Heinze and Schemper, 2002; Zorn, 2005).

Table 2 provides a simple example of quasi-complete separation. The table displays six observations of a binary dependent variable  $y$  and three time dummies:  $\kappa_1$ ,  $\kappa_2$ , and  $\kappa_3$ . As the time index  $t$  shows, the data consists of three durations split into binary data, where the first duration lasts one period, the second lasts two periods, and the third lasts three periods.

<sup>10</sup>Even though Albert and Anderson (1984) only deal with the logit case, cloglog and probit are similarly affected. Data with an ordinal dependent variable can also fall into this classification. See So (1995) for an intuitive and non-technical explanation. See Zorn (2005) for examples of overlap vs separation with a continuous regressor.

<sup>11</sup>For a formal proof of nonexistence in the logit case, see Albert and Anderson (1984).

Consider the first time dummy,  $\kappa_1$ . We see from Table 2 that there is no value of  $\kappa_1$  that perfectly predicts a value of  $y$ .  $\kappa_1 = 1$  is associated with both  $y = 0$  (observations 2 and 4) and  $y = 1$  (observation 1). Similarly,  $\kappa_1 = 0$  is associated with both  $y = 0$  (observation 5) and  $y = 1$  (observations 3 and 6). Examining the second time dummy,  $\kappa_2$ , reveals the same type of overlap between values of  $\kappa_2$  and  $y$ . Now consider  $\kappa_3$ . Here,  $\kappa_3 = 0$  is associated with both  $y = 0$  and  $y = 1$ . However,  $\kappa_3 = 1$  is only associated with  $y = 1$  (observation 6). In other words,  $\kappa_3 = 1$  perfectly predicts that  $y = 1$ . If we were to conduct logistic regression in a package like Stata, the program would inform us that  $\kappa_3$  perfectly predicts  $y = 1$ . Before proceeding with the regression, it would drop  $\kappa_3$  from the analysis, as well as observation 6. If we were to conduct logistic regression using this data, but without checking for separation, we would expect the coefficient on  $\kappa_3$  to be pushed towards  $\infty$ , stopping only due to the default (or user-set) convergence tolerance.

## 2.2 Separation Issues in Practice

Substantive scholars are interested, of course, in the practical consequences of separation on our inferences — in this case, concerning the effect of time on the probability of event occurrence. In that regards, we might ask: (1) Under what conditions is data separation more likely? (2) What effect does separation (and its remedies) have on our inferences concerning time? and, (3) Does separation occur often enough in data analyses that we should really care about it? In later sections comparing time dummies to other techniques for modeling time, we will provide ample evidence of the effect of separation on inferences concerning the hazard. In this section, we focus on the first and third issues.

The conditions under which time dummies tend to produce separation were actually hinted at in Table 2. There, the separation is associated with the dummy  $\kappa_3$  for the latest time period ( $t = 3$ ), which takes the value one in only a single observation. As we will see, time dummies will tend to be associated with separation when the data generating process produces a small proportion of very long durations — and in those cases, the time dummies associated with the separation will tend to be those representing later periods.

In Section 4 we will discuss in more detail a number of Monte Carlo experiments that were conducted to compare the various techniques examined here (e.g., time dummies, splines, cubic polynomial). The Monte Carlo experiments were conducted using various reasonable examples of decreasing, increasing, and non-monotonic hazards.<sup>12</sup> For the

---

<sup>12</sup>The non-monotonic hazard here is referred to as “non-monotonic hazard 2” in Section 4. This hazard

time being, consider Figure 2, which displays the percent of time dummies that needed to be dropped due to separation in the Monte Carlo analyses. Figure 2(a) shows the density of the percent of time dummies dropped under each hazard assumption. In both the decreasing and non-monotonic hazard scenarios, up to 80% of the time dummies are dropped, with averages around 45% and 35% dropped, respectively. The increasing hazard case is not nearly as problematic. Yet, even there, Figure 2(a) shows that up to 25% of time dummies were dropped in some Monte Carlo simulations.

Figure 2(b) provides some insight into the results in Figure 2(a). Each circle in Figure 2(b) displays the percent of time dummies dropped in a particular Monte Carlo experiment as a function of the maximum duration in the data.<sup>13</sup> As the graph shows, there is a fairly straightforward relationship: datasets with larger maximum durations (which require more time dummies, all else equal) tend to drop more time dummies due to separation. However, that’s only part of the story. The other part was seen in Table 2. It is not just the presence of long durations (and therefore more dummies) that is associated with separation, but the presence of long, *unique*, durations. In the increasing hazard scenario, failure occurs gradually over time, but occurs with increasing probability. In contrast, in the decreasing hazard scenario, failure tends to occur rather quickly for most individuals. However, those that survive the higher initial hazard are increasingly likely to have long lives, since the hazard is decreasing. Thus, the decreasing hazard is associated with very long, but rare, durations that create separation problems in binary data with time dummies.

Returning to our first question: when will separation be a problem for time dummies? The preceding analysis shows that decreasing hazards are likely to create problems for time dummies. Non-monotonic hazards may also suffer from similar problems, but that will depend on the form of non-monotonicity (e.g., single-peaked vs u-shaped parabolas). Of course, researchers cannot know beforehand what the true hazard is in their data. However, they can examine their data for possible problems. In that regard, datasets where  $n$  is not overly large (i.e.,  $n < 10000$ ) and where the maximum duration is fairly long (e.g., greater than 15 periods) are likely to suffer from separation.

Finally, do we actually see separation in empirical analyses? A prominent empirical example where quasi-complete separation exists is in the reanalysis of John Oneal and

---

has a sharp rise in failure followed by a gradual decrease over time.

<sup>13</sup>Figure 2b is based on the non-monotonic hazard. The graphs for the increasing hazard and decreasing hazard scenarios look similar and are available upon request from the authors.

Bruce Russett's data in BKT, where 3 of 34 time dummies are dropped, along with 916 observations. A very large  $n$  of 20990 explains why the percentage dropped,  $\approx 0.08$ , is smaller than what would be implied by the graphs in Figure 2 for a decreasing hazard, as the chance that we observe all of the four combinations of the dependent variable and each of the binary regressors should generally increase with the number of observations. A reanalysis of Palmer, London and Regan (2004) reveals much more severe quasi-complete separation issues, with 19 of 39 dummies and 672 of 2975 observations dropped as a result. Clark and Hart (1998) essentially perform a robustness check for the findings of Lemke and Reed (1996) by adding time dummies to the original logit specification. However, the use of time dummies in this case is quite problematic, as quasi-complete separation arises in all of the models. For instance, in Lemke and Reed's model 5, 12 out of 32 total time dummies (37.5%) are perfect predictors of  $Y = 1$ . The consequences of this are not trivial, as it is not possible to plot a significant portion of the hazard plot and 497 of 7031 observations are dropped from the analysis.

Both Monte Carlo results and empirical examples demonstrate that separation is a potentially serious problem that should at minimum give pause to researchers before they utilize time dummies, especially if relatively long durations occur in their data. Efficiency may also be a concern. Fortunately, neither of these are problems if either a cubic polynomial or spline is used.

### **3 Love the Spline, Hate the Spliner**

The second approach to modeling time dependence that BKT advocate is to use splines. In theory, splines are a powerful tool for creating a smooth function of time (or almost any variable that is approximately continuous). In practice, there are many different types of splines (e.g., piecewise linear splines, quadratic splines, natural cubic splines, B-splines), many of which are available as canned packages in Stata, SPSS, or R. An enormous statistical literature has been generated on the subject of splines. In general, we agree with BKT that this is a perfectly reasonable approach to modeling time dependence in binary data.

That said, we saw in Table 1 that most researchers employing splines never actually discuss the effect of time. As we have noted, we suspect this is because splines are a

relatively unknown (and somewhat complicated) technique for most political scientists.<sup>14</sup> Because of that, researchers have tended to use BKT’s software with its default parameter settings (specifically, concerning knot placement, more about which later). These default parameter values may be inappropriate in many cases and result in biased inferences concerning the hazard.

Explaining the details of all variants of splines is beyond the scope of this paper.<sup>15</sup> However, some intuition concerning splines will be helpful for at least two reasons. First, BKT do not provide enough detail for users to employ splines either in the regression analysis or in plotting the hazard. For those who want to model time dependence using splines, this section will help researchers understand, implement, and interpret splines in modeling time dependence. We introduce basic types of splines, relate them to the BKT approach for modeling temporal interdependence, and address issues with splines that practitioners should be aware of. We also provide (in the appendix) an R program that researchers can modify and use to conduct their analyses and then plot the hazards. Secondly, the (much simpler) cubic polynomial technique we introduce in Section 4 is related to cubic splines. Although the technique performs just as well as splines in most reasonable situations, it will be useful for researchers to understand the trade-offs between the two approaches.

### 3.1 Splines

For our purposes, we can think of a spline as a smoother — a function that allows us to smooth the relationship between two variables, say a dependent variable  $y$  and time  $t$ . Most splines allow one to specify points in  $t$  where the relationship with  $y$  changes substantially. Those points are referred to as “knots.” Fewer knots will lead to a smoother relationship, but may miss important changes in the relationship. Specifying more knots allows for more changes to be modeled. On the other hand, the greater the number of knots, the less smooth the relationship. Moreover, the spline may pick up on idiosyncratic changes in the relationship, not general trends.

A key component to all splines is that they generate a set of vectors that are a function

---

<sup>14</sup>Indeed, many authors are quite upfront about viewing splines as an opaque method that controls for a statistical nuisance. One author refers to them as “so-called cubic splines” that are used “to control for any temporal correlation” (Dorussen, 2006, 95). Another set of authors criticize the standard approach as “hiding” important linkages “in the peace-year spline variable” (Goertz, Jones and Diehl, 2005, 747).

<sup>15</sup>We instead refer readers to Keele (2008) for a recent introduction to splines by a political scientist.

of the independent variable (here time  $t$ ) and associated with the knots. These vectors are referred to as basis vectors and the set of vectors is referred to as a basis matrix.<sup>16</sup> A basis is a set of linearly independent vectors, the linear combination of which can span an entire vector space (i.e., reach any point in that space). Most importantly for substantive researchers, these basis vectors can be included as regressors in our statistical analysis, allowing us to estimate a wide array of smooth functions of time.

To illustrate this, let us consider two of the simplest splines: a piecewise linear spline and a simple cubic spline. Suppose we believe the probability of some event (e.g., war) in observation  $i$  is explained by a set of regressors  $x_{i,t}$  and some smooth function of time  $s(t_i)$ . We can specify the logit equation for this situation as

$$\Pr(y_{i,t} = 1) = \frac{1}{1 + \exp[-(x_{i,t}\beta + s(t_i))].}$$

For the piecewise linear spline,  $s(t_i)$  takes the form

$$s(t_i) = \alpha_1 t_i + \sum_{k=1}^K \alpha_{1k} (t_i - \eta_k)_+ \quad (3)$$

where the function  $(t_i - \eta_k)_+$  returns the difference between  $t_i$  and  $\eta_k$  when it is positive, but equals zero otherwise. The  $\eta_k$  are the  $k$  user-specified knots. If one were to use a piecewise linear spline (something we do not necessarily recommend), one would include  $t_i$  and the  $(t_i - \eta_k)_+$  vectors as regressors. The  $\beta$  and  $\alpha$  coefficients would then be estimated in the logistic regression.<sup>17</sup>

The simple cubic spline is conceptually very similar:

$$s(t_i) = \alpha_1 t_i + \alpha_2 t_i^2 + \alpha_3 t_i^3 + \sum_{k=1}^K \alpha_{1k} [(t_i - \eta_k)_+]^3. \quad (4)$$

The main differences here are the inclusion of the  $t_i$  polynomials, as well as the cubed  $(t_i - \eta_k)_+$  terms for each knot. Again, if one were to use a simple cubic spline, one would include  $t_i$ ,  $t_i^2$ , and  $t_i^3$ , and the  $[(t_i - \eta_k)_+]^3$  vectors as regressors, and then estimate the  $\beta$  and  $\alpha$  coefficients via logistic regression.

<sup>16</sup>Suppose we have some vector space  $S$  such that it is possible to express every vector in  $S$  as a linear combination of  $k$  ( $x_1, x_2, \dots, x_k$ ) vectors. Then, if the  $k$  vectors are linearly independent, the set of  $k$  vectors forms a basis for  $S$ . See Searle (1982) for more details.

<sup>17</sup>An interesting special case of the piecewise linear spline is when a knot is placed at each time point in the data. In this case, the piecewise spline is equivalent to time dummies, and the estimated results are identical.

Commonly used splines, such as the natural cubic spline employed by Beck, Katz and Tucker (1998), are just more complicated versions of equation 4, but which have certain nice properties. For example, natural cubic splines are linear beyond their boundary knots. B-splines are similar to natural cubic splines, but have the nice properties that the basis vectors will be orthogonal to each other (i.e., no multicollinearity issues) and vary between 0 and 1 (i.e., no numerical instability issues). For the practitioner, the steps to employ these splines in a regression are the same as above: (1) choose knot locations, (2) use the chosen spline procedure (with selected knots) on the time variable  $t_i$  to generate basis vectors, (3) include those basis vectors as regressors, and (4) estimate coefficients for the substantive regressors and the spline basis vectors.

Estimating the  $\beta$  and  $\alpha$  coefficients is, of course, only the first step in substantive analysis. Researchers may interpret the effects of substantive regressors in the usual ways (e.g., first difference, plots of fitted values, etc). Rather than simply including the basis vectors produced by splines as regressors and stopping there, researchers should think theoretically about whether and why temporal dependence is present in their data. For example, is the temporal dependence due to omitted variables? Or, does the temporal dependence reflect a part of the process that needs to be modeled?

Although BKT provide Stata code for generating splined time basis vectors, no directions are given concerning how to interpret the results. Table 1 showed the result: almost no one using splines interprets the results. It is not difficult to imagine why this would be the case. It is one thing to insert basis vectors into a regression equation and to estimate the coefficients associated with the basis vectors. However, how does one then interpret the estimated results for the basis vectors?

It is important to note first that the estimates for the individual spline basis vectors are not substantively interesting. For example, we are not interested in the individual effect of the  $t_i^2$  variable or of the  $[(t_i - \eta_2)_+]^3$  vector in Equation 4. Rather, we want to interpret the effect of time  $t$  on the estimated  $\Pr(y_{i,t} = 1)$ . The problem for many substantive researchers, we suspect, is that the spline-based regression equation is in terms of the spline basis vectors, not in terms of  $t$ .

Assuming one has already conducted logistic regression and included the spline basis vectors as regressors, the steps for plotting the estimated hazard are relatively straightforward. First, construct a new time vector  $\tilde{t} = \{1, 2, 3, \dots, \max(t)\}$ . This will be used in the next step and will serve as the  $x$ -axis for the plot. Next, apply the spline function

that was used for the data analysis *in exactly the same way* (i.e., same number and location of knots) to the  $\tilde{t}$  vector. This will provide basis vectors for  $\tilde{t}$  that correspond to those used in the regression. The ordering of the observations will also correspond to the ordering of time in  $\tilde{t}$ . Then, using the logistic regression’s parameter estimates, calculate  $\Pr(y_{i,t} = 1)$  for each row in the newly generated ( $\tilde{t}$ ) basis vectors, substituting those basis vectors into their corresponding locations in the regression equation, while holding all other (non-spline) variables constant at some value (e.g., their means or medians). The estimated probabilities will be ordered according to the ordering in  $\tilde{t}$ . The researcher then needs only plot the estimated  $\Pr(y_{i,t} = 1)$  with  $\tilde{t}$  along the  $x$ -axis. In the Appendix, we provide an R program that demonstrates the above steps.

### 3.2 Knot Selection

Although we have ignored it to this point, one of the most important aspects of implementing splines is appropriately selecting the knots  $\eta_k$ . The number of knots determines the number of basis vectors that are included in subsequent regression analysis. The locations of the knots determine where the relationship between the dependent variable and  $t$  is allowed to change significantly. How, then, should a researcher select the knot locations?

When the dependent and independent variables are both continuous, knots are often selected using an “ocular” diagnostic — one simply examines a scatterplot of  $y$  vs  $t$  and chooses knot locations where the relationship appears to change. As Ruppert, Wand and Carroll (2003, 57–72) note, this strategy that can be quite effective. However, with a binary dependent variable, knot selection is more difficult. A plot of the binary data versus  $t$  will generally not provide any guidance concerning where to place knots. Moreover, the quantity of interest here,  $\Pr(y_{i,t} = 1)$ , must be estimated in order to plot it as a function of  $t$ . We do not know what it looks like before estimating it — and what it looks like after estimating it will depend on the knots we chose.

In this case, substantive researchers may want to base knot placement on (1) theory and/or (2) an empirical selection criterion. Unfortunately, political theory is typically not very specific concerning knot placement. The most we can usually hope for is a sense of whether the theory implies an increasing, decreasing, or a non-monotonic failure rate. However, that is of limited use in choosing knot locations. Given the difficulty in choosing theoretically informed knots, most researchers, such as Schultz (2001, 270–271), have employed the same knots as in Beck, Katz and Tucker (1998) (i.e., knots at  $t = 1, 4,$

and 7). Others, such as Senese (2005), simply provide no discussion of knot selection at all.

One alternative is to base knot selection on some criterion associated with model fit. Beck, Katz and Tucker (1998, 1279) choose knots via “a sequence of F-tests” — basically comparing the fit for various knot combinations. It is important to note that the BKT knots are based on an empirical selection criterion tailored to their specific example (Oneal and Russett (1997)’s analysis of trade and conflict). There is no reason to believe that these knots will be appropriate for other substantive issues or data. Depending on the number of time intervals in the data and the shape of the underlying hazard, knots at  $t = 1, 4, 7$  may actually produce a biased estimate of the hazard. We provide an example of this in the next section.

Finally, automated smoothing methods are an extension of this empirical selection approach. Generalized cross validation (GCV) and a random effects-based techniques essentially choose many knots, but penalize the influence of knots to balance smoothness versus over-fitting. Although a more detailed discussion of automated smoothing techniques is beyond the scope of this paper, we have included examples in the plots of the Monte Carlo analyses in Section 4.<sup>18</sup> Our intention is not to rule out the use of either splines or automated smoothing techniques. However, the issues discussed in this section, combined with the experimental and empirical evidence we provide in the next section, suggest that a simpler cubic polynomial approximation performs just as well as splines in most substantive settings, but without the additional complexity.

## 4 Time Cubed

Having discussed the technical details of time dummies and splines, we now recommend an alternative method for modeling time dependence. In fact, our recommendation is almost embarrassingly simple: include  $t$ ,  $t^2$ , and  $t^3$  as regressors. To make this concrete, suppose a researcher with regressors  $x_{i,t}$  wanted to conduct logistic regression, control for temporal dependence, and interpret the effect of time on  $\Pr(y_{i,t} = 1)$ . Using our approach, her logit equation would take the form

$$\Pr(y_{i,t} = 1) = \frac{1}{1 + \exp [-(x_{i,t}\beta + \alpha_1 t_i + \alpha_2 t_i^2 + \alpha_3 t_i^3)]} \quad (5)$$

---

<sup>18</sup>Several forms of GCV can be implemented using the `mgcv` package in R. See Ruppert, Wand and Carroll (2003) and Keele (2008) for details on automated smoothing methods.

where  $s(t_i) = \alpha_1 t_i + \alpha_2 t_i^2 + \alpha_3 t_i^3$  is a cubic polynomial approximation to the hazard. Notice that the cubic polynomial is a special case of the simple cubic spline in Equation 4 — one with no knot terms.

Why use a *cubic* polynomial? In principle, any order of polynomial could be chosen. As with splines, the polynomial approach can be taken to a higher order (e.g.,  $t^4$ ,  $t^5$ , etc). Generally, polynomials of odd order are preferable to polynomials of even order. Polynomials of odd order asymptotically have a smaller mean-squared error than polynomials of even order (see Fox (2000) for a simple illustration of this). We recommend a cubic polynomial because it will capture any hazard shape that is recovered by commonly estimated parametric duration models (e.g., Weibull, log-logistic, log-normal) and typically seen in semi-parametric models such as the Cox proportional hazard model. At the same time, the cubic polynomial avoids the over-fitting associated with higher order polynomials (e.g., plots with “kinks” that are sensitive to a small proportion of observations). That said, if there are strong theoretical reasons for using a higher order polynomial, then the polynomial approach is easily extended.

Using this technique, interpreting the effect of time is straightforward. Indeed, it is no different than any regression that includes quadratic or cubic variables. As with splines, we are interested in plotting the fitted  $\Pr(y_{i,t} = 1)$ , rather than interpreting the individual polynomial terms. Having estimated the model, the researcher simply creates a new time vector  $\tilde{t} = \{1, 2, 3, \dots, \max(t)\}$ , along with its square and cube,  $\tilde{t}^2$  and  $\tilde{t}^3$ , respectively. The fitted values for  $\Pr(y_{i,t} = 1)$  are calculated by inserting these into the estimated regression equation, holding all other variables constant at some interesting profile of values (most commonly means). The estimated  $\Pr(y_{i,t} = 1)$  is then plotted on the y-axis with  $\tilde{t}$  on the x-axis.

Finally, although we have found the cubic polynomial method to be relatively insensitive to computational issues in practice, researchers should be aware of two potential issues. First, it is generally well known that the  $t$ ,  $t^2$ , and  $t^3$  variables will be highly correlated. For large data sets, this is almost never a problem. Moreover, we have found no evidence that it presents any problems in numerous Monte Carlo experiments and empirical reanalyses. Nevertheless, researchers suspecting multicollinearity may want to de-mean  $t$  before squaring and cubing it. This will reduce (although not completely eliminate) collinearity. Secondly, numerical instability is always an issue in any maximum likelihood estimation. One common source of numerical instability is when one of the regressors is three or four

orders of magnitude larger than all the other regressors. Using cubic polynomials, instability could result from large differences in magnitude between  $t^3$  and other regressors, given that  $t^3$  can be quite large depending on the maximum value of  $t$ . For example, if the maximum duration is  $t = 25$  then  $t^3$  varies from 0 to 15625. The solution here is no different than what common practice should already be. Simply examine the range of all the data and  $t$ ,  $t^2$ , and  $t^3$  and rescale variables as necessary via division by some multiple of 10. We have found utilizing  $\frac{t}{100}$  and its square and cube generally works quite well. Using  $t$ ,  $t^2$ , and  $\frac{t^3}{1000}$  also works well. In either case, users should remember to use the appropriate scaling when plotting the estimated  $Pr(y_{i,t} = 1)$ .

## 5 Monte Carlo Comparison

In this section, we provide a comparison of the various techniques via Monte Carlo analysis. In order to compare the various methods, we will assume that the data generating process is logit with time dummies.<sup>19</sup> As previously discussed, employing time dummies as the data generating process is consistent with an underlying Cox framework. Time dummies also allow us to create hazards of just about any functional form. Finally, using time dummies to generate the data does not induce any of the problems associated with their use in estimation. Thus, we will assume the data generating process takes the form

$$Pr(y_{i,t} = 1) = \frac{1}{1 + \exp(-(x_{i,t}\beta + \kappa_{i,t}\gamma))} \quad (6)$$

where  $x_{i,t}$  consists of a constant and a regressor that is uniformly drawn between -2 and 2, and all  $\beta$ 's are equal to one. The  $\kappa_{i,t}$  are the time dummies. The shape of the hazard is therefore determined by  $\gamma$ 's, which are the coefficients on each of the time dummies. We conduct Monte Carlo experiments for constant, decreasing, increasing, and two different non-monotonic hazards. We run 10,000 Monte Carlo iterations with samples of  $n = 2000$  for each iteration. In each iteration for each hazard shape, we estimate logistic regressions with time dummies, cubic B-splines, automatic smoothing via GCV, and a cubic polynomial. To assess how well splines perform as implemented in the discipline we use knots at  $t = 1, 4, 7$  for the B-spline estimation. Figure 3 provides graphical illustrations that demonstrate how well each method performs on average in recovering the true underlying hazard. In each graph in Figure 3, the true hazard is depicted by the thick gray line; the cubic polynomial approximation by a solid black line; the automated

<sup>19</sup>We observed virtually identical results when the data generating process is cloglog with time dummies.

smoothing spline by a dashed line; the B-spline (with knots at  $t = 1, 4, 7$ ) by a dotted line; and the model estimated using time dummies by the open circles.

Consider first Figure 3(a), where the true hazard is decreasing in time. On average, hazards estimated using the cubic polynomial or either of the spline variants closely match the true hazard. Time dummies, on the other hand, perform considerably worse. We noted in Section 2.2 that a decreasing hazard will tend to be associated with a higher incidence of data separation. We see in Figure 3(a) the effect of those separation problems. Most problematic, we would on average infer that a non-monotonic hazard existed, since the time dummies plot begins increasing sharply at about  $t = 20$ . Secondly, even with the large number of monte carlo iterations, separation prohibits estimating effects for the last few periods in the data. Third, although not shown here, the 95% confidence interval for the time dummies plot does not contain the true hazard from approximately  $t = 28$  on.<sup>20</sup>

Figure 3(b) displays the results when the true hazard is increasing in time. In this case, the cubic polynomial and the auto-smoothing spline both perform very well, with the auto-smoothing spline having a slightly closer fit. Time dummies also perform reasonably well here. We noted in Section 2.2 that separation is likely to be less of an issue for increasing hazards. That is certainly reflected in Figure 3(b). The only issue for time dummies in this particular case is that it cannot estimate effects past about  $t = 28$ . The method that performs most poorly here is the B-spline with knots set arbitrarily at  $t = 1, 4$  and  $7$ . The poor fit is entirely due to the knot selection (relative to the true hazard). Notice the spline fits very well up to about  $t = 12$ . However, the time periods extend along the x-axis to about  $t = 35$ . Most importantly, much of the curvature in the true hazard exists around  $t = 20$ . By placing the last knot at  $t = 7$ , there is no way for the spline to adjust its curvature at later time points. Thus, researchers should only use a spline with knots at  $t = 1, 4, 7$  when (1) the longest time period is around 10 or 12, or (2) they truly believe all the curvature in the hazard occurs earlier in time rather than later.

Figures 3(c) and 3(d) show the results for two different non-monotonic hazards. The u-shaped hazard in Figure 3(c) reflects a situation where the failure rate is initially moderate, decreases to a very low level, but then eventually rises to a very high level.<sup>21</sup> As in the two preceding cases, the cubic polynomial performs very well. The auto-smoothing spline fares reasonably well, as do the time dummies. For the time dummies, the increasing

<sup>20</sup>Graphs of confidence intervals are available upon request from the authors.

<sup>21</sup>Life expectancy in Western countries takes a form similar to this hazard.

hazard later in time prevent the really problematic separation issues seen in Figure 3(a). However, notice the (somewhat disturbing) fact that time dummies produce a u-shaped hazard in both Figure 3(a) (where it is incorrect) and 3(c) (where it is correct). Finally, the B-spline with knots at  $t = 1, 4, 7$  fits well up to approximately  $t = 20$ , but diverges thereafter. The reason is exactly the same as in the previous graph: placing the last knot at  $t = 7$  constrains it from adjusting to curvature later in time (e.g., around  $t = 25$ ).

The non-monotonic hazard in Figure 3(d) is produced from the log-logistic distribution and captures situations in which the probability of observing the event of interest within a group initially increases sharply, but then decreases with time. The cubic polynomial and spline variants generally perform well in this scenario. Because most of the changes in curvature occur prior to  $t = 12$ , the B-spline fits that part of the hazard quite well, but diverges a bit more after  $t = 15$ . The cubic polynomial tries to fit the hazard overall. Because of that, it does not fit the true hazard quite as well as the B-spline in the region up to  $t = 12$ . However, it tends to have a better fit from about  $t = 15$  on. Time dummies perform very well up to about  $t = 12$ . After that, time dummies suffer from the same separation problems as in the decreasing hazard plot, and they diverge greatly after  $t = 15$ .

In sum, the cubic polynomial approximation — simply including  $t$ ,  $t^2$ , and  $t^3$  as regressors — tends to recover the true hazard on average and for a wide variety of hazard shapes. The Monte Carlo results strongly suggest that time dummies should almost never be used. This is a particularly interesting result, given that time dummies were used as the data generating process. A spline (whether B-spline or natural cubic spline) with knots at  $t = 1, 4, 7$  will perform poorly unless most of the changes in the relationship occur with the region  $t \in [0, 12]$ . Therefore, if a researcher wants to implement a spline-based technique, but is not confident of knot placement, she should opt for an auto-smoothing spline.

## 6 An Application to Associational Formation in the United States

In order to demonstrate empirically the usefulness of the cubic polynomial approach, we turn to Crowley and Skocpol (2001)'s analysis of why, when, and how membership associations (such as the Young Men's Christian Association, or YMCA) formed in the United States.<sup>22</sup> Crowley and Skocpol (2001) focus on two competing perspectives. The first per-

<sup>22</sup>We should note that we have replicated a number of other analyses: e.g., the reanalysis of Oneal and Russett by Beck, Katz and Tucker (1998), as well as the analysis in Palmer, London and Regan (2004). In

spective, popular among historians, is that socio-economic modernization and high levels of immigration were the main catalysts of associational development (Berthoff, 1971; Keller, 1977; Wiebe, 1967). In contrast, Crowley & Skocpol’s competing historical-institutionalist perspective argues that the U.S. Civil War played a pivotal role in associational development (Schlesinger, 1944; Skocpol, Ganz and Munson, 2000; Crowley and Skocpol, 2001). In their view, since the U.S. government lacked a large standing army or well-developed bureaucracy prior to the Civil War, voluntary federations were assembled across the states to aid in raising an army (McPherson, 1988). After Northern victory, civic associational structures built to support the war effort remained in place or were replicated by the founders of new civic associations. Thus, the historical-institutionalist account implies that associational development was propelled by the after-effects of association building and development during the U.S. Civil War.<sup>23</sup> Crowley and Skocpol (2001, 814, 820) explicitly states that the passage of time is an important component of their theory.

To assess the support for the two perspectives, Crowley and Skocpol (2001) construct a new data set that tracks the establishment of state-level chapters for twenty-one organizations in the forty-eight continental states over the period 1860-1929. Each observation, therefore, indexes by state, organization, and decade. The binary dependent variable represents whether a state-level branch of a particular membership federation was established in a given decade. States become “at risk” of chapter formation for a particular organization once that organization has been established nationally in at least one other state. For a given state-organization observation, the duration component reflects the number of decades that state has been at risk for local chapter formation by that organization. Once a chapter has formed, that state-organization pair exits the sample. Independent variables represent various measures of the modernization and institutionalist perspectives. Crowley and Skocpol (2001) estimate a logistic regression with time dummies to determine the key determinants of associational formation.

Table 3 displays a replication of the Crowley and Skocpol (2001) regression. Three versions are presented: one with time dummies, one using a cubic polynomial, and the last using a spline with automatic smoothing.<sup>24</sup> As Table 3 shows, the substantive results

---

all cases, we find that a cubic polynomial performs just as well as splines with automatic knot placement. Results are available upon request from the authors.

<sup>23</sup>For a complete account, see Crowley and Skocpol (2001, 814–816) and the sources cited therein.

<sup>24</sup>Although some of the numbers in the “time dummies” column of Table 3 differ slightly from those in Crowley and Skocpol, the results are exactly the same in terms of statistical significance and substantive effect. The numerical differences are due only to choices concerning which time dummy to drop and rescaling of certain regressors. For example, Crowley and Skocpol drops the sixth time dummy, while we

are generally the same regardless of the technique used to model temporal dependence. The only difference here is that the variable for percent protestant is statistically significant at  $p \approx .053$  in the time dummies model, but at  $p \approx .01$  in the cubic polynomial and spline models.

## 6.1 Associational Formation over Time

While Crowley and Skocpol (2001) nicely analyzes how associational development was influenced by the Civil War, it has less to say about *when* these effects materialized across time. Despite including time dummies as “temporal controls,” there is no interpretation of the hazard. As it turns out, both the socio-economic modernization and historical-institutionalist perspectives have implications concerning associational formation over time. Assuming that states do not decrease in modernization over this period, then if the socio-economic modernization account is correct, we should expect to see the probability of state-level formation increase (or at least not decrease) with time. On the other hand, if the historical-institutionalist account is correct, then we should expect to see an initially large post-Civil War effect, which then decreases over time.

Figure 4 plots the probability of state-chapter formation as a function of time, holding all other variables at their means (continuous variables) or medians (discrete variables). As the figure shows, the probability of state chapter formation initially grows to a relatively high level in the thirty years after the first chapter is established, but then falls to a very low level over the remaining four decades. Because most of the organizations — all but four — established either national centers or their first state chapters prior to the turn of the century (see Crowley and Skocpol (2001, 817)), the hazard in Figure 4 appears to be much more consistent with the Civil War-based, historical-institutionalist account.

Figure 4 is a starting point for interpreting the role of temporal dependence in associational formation. However, it only tells part of the story. One of the key Civil War variables, and the most substantively significant, codes for the average pension received by Union Civil War pensioners in a given state. As (Crowley and Skocpol, 2001, 818) note, high pensions tended to be associated with military and political elites, as well as the seriously wounded, who were themselves often highly engaged in the community. Suppose we wanted to assess the hazard at different levels of pension pay in states.

---

drop the first. We also divided several of the variables by a multiple of 10 to ensure all regressors were of roughly the same magnitude.

Figure 5(a) plots the probability of state chapter formation as a function of time and as a function of average pension pay, based on the cubic polynomial model in Table 3. As the figure shows, the probability of chapter formation takes the same shape regardless of average pension pay: it increases over the first three decades and then falls over the next four. Thus, examining the hazard for one level of pension pay tells us nothing different than does another level of pension pay. The reason for this is that the hazard in Figure 5(a) is based on a proportional hazards model. As Box-Steffensmeier and Zorn (2001) note, nearly all duration models commonly used by political scientists assume proportional hazards.<sup>25</sup> Indeed, *all* of the models in Table 3 are proportional hazard models. When the proportional hazards assumption is incorrect, our parameter estimates and standard errors may be substantially biased. Moreover, there are many instances where our theory suggests the hazards are *nonproportional*.

Fortunately, modelling nonproportional hazards is relatively easy when using  $t$ ,  $t^2$ , and  $t^3$ . To do so, we recommend simply interacting the regressor(s) of interest with  $t$ ,  $t^2$ , and  $t^3$  (or appropriately scaled versions thereof).<sup>26</sup> Testing for nonproportionality is also straightforward. The user first estimates the unrestricted model with both the time variables (e.g.,  $t$ ,  $t^2$ ,  $t^3$ ) and the variable-time interaction terms (e.g.,  $xt$ ,  $xt^2$ ,  $xt^3$ ). Then the same model is estimated, including the time variables, but omitting the interaction terms (i.e., restricting their coefficients to zero). The test for nonproportionality is then a likelihood ratio test based on the unrestricted and restricted models.

Returning to associational formation, if the historical-institutionalist perspective is correct, then we should expect that the hazard will differ for states with high average pensions versus those with low average pensions. In particular, states with high average Union pensions (and, thus, more federation-oriented community leaders) will have a higher probability of immediately forming state-level chapters. On the other hand, states with low average Union pensions will need time for those ideas (and/or pensioners) to take root.

To examine the support for this idea, we only need conduct one more regression. The cubic polynomial regression in Table 3 is our restricted model. The second regression, our unrestricted model, is exactly the same, but also includes the interaction of the average

---

<sup>25</sup>For a technical discussion of the differences between proportional and nonproportional hazard models, see Kalbfleisch and Prentice (2002), Box-Steffensmeier and Jones (2004, 131–137), or Box-Steffensmeier and Zorn (2001).

<sup>26</sup>Nonproportional hazards can also be modeled in a similar way using time dummies or splines. However, we do not recommend doing so with time dummies, as this literally multiplies the potential problems already identified. Thanks are due to Hein Goemans for suggestions concerning nonproportional hazards.

pension dollars variable with each of the  $t$ ,  $t^2$ , and  $t^3$  variables. A likelihood ratio test rejects the restrictions at well below the 0.05 level. Thus, the nonproportional hazard model appears warranted in this case.<sup>27</sup>

The graph in Figure 5(b), plots the estimated probability that a federated voluntary association forms as we vary both time and the average pension dollars per pensioner in each state. Compared to the proportional hazard plot in Figure 5(a), the results are striking. First, consider the effect of pension dollars on the hazard. For states with high average pension per pensioner (e.g., 200–250 per pensioner), there is an immediately high probability of chapter formation (.52), dropping fairly linearly to almost zero after seven decades. On other hand, the hazard is nonmonotonic in states with the lowest average pay (0–50 per pensioner), starting at .17 probability of chapter formation, and needing a couple of decades before the probability reaches approximately .46, after which it too steadily declines. This relationship indicates that organizations took off very fast where the Union armies had a large postbellum footprint and were generally followed several decades later by states where the organizational influence of the Union army was either initially negligible or took time to spread into those areas.

Figure 5(b) also provides insight into the role of pension pay that we would otherwise not have found in the proportional hazards model. First consider the interpretation of average pension pay from either Table 3 or Figure 5(a) (which simply reflects the former). In the proportional hazards model, average pension pay has a positive effect regardless of time. However, consider the more nuanced interpretation in the nonproportional hazards Figure 5(b). In the first decade ( $Time = 0$ ) after a federation is established (which is relatively soon after the Civil War ends in most cases), average pension pay has a very large and positive effect. The difference in probability between highest pension areas versus lowest pension areas is .36. In contrast, at the end of seven decades after national federation formation, the effect of pension pay is *negative*: the difference between the highest and lowest pension areas is  $-.11$ . Why might that be? We suspect it reflects the higher initial rate of chapter formation in high pension areas. After seventy years, most have already formed chapters. The lowest pension states lag behind, so there are still chapters to be formed.

In sum, while Crowley and Skocpol were correct about the positive impact of Union

---

<sup>27</sup>Although not shown, the parameter estimates on the main regressors remain similar to those in the restricted model. Results available from the authors.

armies, they did not fully explore the temporal dimension of the relationship. A more nuanced analysis, including one that models nonproportional hazards, provides even stronger support for the historical-institutionalist perspective.

## 7 Conclusion

Our goal in this article has been to build upon the important contributions made in Beck, Katz, and Tucker (1998). First and foremost, BKT makes an important methodological contribution in demonstrating that BTSCS data is a form of grouped duration data. This observation, coupled with recommendations for how to deal with temporal dependence has generally improved the quality of empirical research analyzing temporally dependent binary data.

That being said, we have identified a number of issues concerning the techniques themselves or their implementation by substantive researchers. We have demonstrated that time dummies often suffer from complete and quasi-complete separation, as well as inefficiency issues. Our monte carlo analysis strongly suggests that time dummies should almost never be used when there are a large number of time periods. Moreover, the analysis also demonstrates that time dummies perform worse than other techniques even when the data generating process employs time dummies.

Splines are a different issue. Splines can be a useful technique for modeling temporal independence. However, most substantive researchers do not appear to understand splines, resulting in hazards not being interpreted in empirical analysis. Further evidence for this supposition is the fact that most researchers use BKT’s application-specific knots of 1, 4, and 7, regardless of whether this knot selection is appropriate for the researcher’s own application or not. We believe there is more to “taking time seriously” than just adding time regressors as “controls.” We argue that researchers should also plot and interpret the hazard rather than treat temporal dependence as a nuisance.

The method we recommend in this article is to include a cubic polynomial approximation —  $t$ ,  $t^2$ , and  $t^3$  — as regressors. This technique is simple to implement and to interpret. It can also model a wide variety of hazard shapes. Our monte carlo analysis (and numerous empirical replications) suggest it performs very well in practice relative to time dummies and splines. Additionally, when theory suggests that the hazard may be nonproportional, modeling nonproportional hazards is straightforward using the cubic

polynomial, as is testing for nonproportionality. We demonstrate the ease of implementing nonproportional hazards in our replication and extension of Crowley and Skocpol (2001)'s work on federated voluntary associational formation in the postbellum United States. In fact, our more nuanced analysis provides even more support for their theory.

## A Example R Programs

The following two sections provide R code demonstrating how to estimate logit models that smooth time using (1) cubic splines and (2) a cubic polynomial approximation (i.e.,  $t$ ,  $t^2$ , and  $t^3$ ). The code assumes the dataset has already been loaded into R.  $y$  is the dependent variable;  $x_1$  and  $x_2$  are the independent variables; and  $tim$  is the variable denoting the time since the last event.

### A.1 Logit with Cubic Splined Time

```
# Load package with spline procedures
library(splines)

# Generate spline basis vectors using knots=1,4,7.
# Note: these are simply examples used in BKT and may not be
# appropriate for your application.
k <- c(1,4,7)
basis <- ns(tim,knots=k)

# Estimate logit model with cubic splined time.
# Must omit either constant or one of spline base vectors.
# Here: omitting constant.
fit <- glm(y~x1+x2+basis-1,family=binomial(link="logit"))

# To calculate fitted values, first create sequence of time values
# over entire range. Assumes time takes only integer values.
timseq <- 0:max(tim)

# Generate basis for new time sequence and then append means of x1 & x2.
# Must use same knots as those in original data analysis.
# Note: Variables need to be in the same order as estimated betas.
newbasis <- ns(timseq,knots=k)
newX <- cbind(mean(x1),mean(x2),newbasis[,])

# Use newX and estimated betas for XB
Bhat <- fit$coefficients
XBhat <- newX%*%Bhat

# Calculate logit fitted values
prob <- 1/(1+exp(-XBhat))

# Plot estimated Pr(Y=1|t)
plot(timseq,prob,type="n",xlab="t",ylab="Pr(Y=1|t)",lty=1,lwd=2)
```

## A.2 Logit with $t$ , $t^2$ , $t^3$

```
# Create t^2 and t^3 variables
tim2 <- tim^2
tim3 <- tim^3

# Estimate logit with cubic polynomial
fit <- glm(y~x1+x2+tim+tim2+tim3,family=binomial(link="logit"))

# To calculate fitted values, first create sequence of time values
# over entire range.
timseq <- 0:max(tim)

# Create X matrix by appending 1 for constant, means of x1 & x2,
# and then time sequence, squared, and cubed.
# Note: These need to be in the same order as estimated betas.
newX <- cbind(1,mean(x1),mean(x2),timseq,timseq^2,timseq^3)

# Use newX and estimated betas for XB
Bhat <- fit$coefficients
XBhat <- newX%*%Bhat

# Calculate logit fitted values
prob <- 1/(1+exp(-XBhat))

# Plot estimated Pr(Y=1|t)
plot(timseq,prob,type="n",xlab="t",ylab="Pr(Y=1|t)",lty=1,lwd=2)
```

## References

- Albert, A. and J.A. Anderson. 1984. "On the Existence of Maximum Likelihood Estimates in Logistic Regression Models." *Biometrika* 71(1):1–10.
- Alt, James E., Gary King and Curtis S. Signorino. 2001. "Aggregation Among Binary, Count, and Duration Models: Estimating the Same Quantities from Different Levels of Data." *Political Analysis* 9(1):1–24.
- Arceneaux, Kevin. 2003. "The Conditional Impact of Blame Attribution on the Relationship Between Economic Adversity and Turnout." *Political Research Quarterly* 56(1):67–75.
- Balla, Steven J. 2001. "Interstate Professional Associations and the Diffusion of Policy Innovations." *American Politics Research* 29(3):221–245.
- Barbieri, Katherine and Rafael Reuveny. 2005. "Economic Globalization and Civil War." *Journal of Politics* 67(4):1228–1247.
- Bearce, David H. and Sawa Omori. 2005. "How Do Commercial Institutions Promote Peace?" *Journal of Peace Research* 42(6):659–678.
- Beck, Nathaniel, Jonathan N. 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):1260–1288.
- Beck, Neal, Gary King and Langche C. Zeng. 2000. "Improving Quantitative Studies of International Conflict: A Conjecture." *American Political Science Review* 94(1):21–35.
- Bennett, D. Scott and Allan C. Stam. 2000*a*. "Research Design and Estimator Choices in the Analysis of Interstate Dyads - When Decisions Matter." *Journal of Conflict Resolution* 44(5):653–685.
- Bennett, D. Scott and Allan C. Stam. 2000*b*. "A Universal Test of an Expected Utility Theory of War." *International Studies Quarterly* 44(3):451–480.
- Bennett, D. Scott and Matthew C. Rupert. 2003. "Comparing Measures of Political Similarity - An Empirical Comparison of S Versus Tau-b in the Study of International Conflict." *Journal of Conflict Resolution* 47(3):367–393.

- Benson, Michelle. 2005. "The Relevance of Politically Relevant Dyads in the Study of Interdependence and Dyadic Disputes." *Conflict Management and Peace Science* 22(2):113–133.
- Bernard, Michael, Christopher Reenock and Timothy Nordstrom. 2003. "Economic Performance and Survival in New Democracies - Is There a Honeymoon Effect?" *Comparative Political Studies* 36(4):404–431.
- Berthoff, Rowland. 1971. *An Unsettled People: Social Order and Disorder in American History*. New York, NY: Harper and Row.
- Besancon, Marie. 2005. "Relative Resources: Inequality in Ethnic Wars, Revolutions, and Genocides." *Journal of Peace Research* 42(4):393–415.
- Boehmer, Charles R. and David Sobek. 2005. "Violent Adolescence: State Development and the Propensity for Militarized Interstate Conflict." *Journal of Peace Research* 42(1):5–26.
- Box-Steffensmeier, Janet and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*. Cambridge, UK: Cambridge University Press.
- Box-Steffensmeier, Janet and Christopher J.W. Zorn. 2001. "Duration Models and Proportional Hazards in Political Science." *American Journal of Political Science* 45(4):972–988.
- Braithwaite, Alex. 2005. "Location, Location, Location ... Identifying Hot Spots of International Conflict." *International Interactions* 31(3):251–273.
- Brinks, Daniel and Michael Coppedge. 2006. "Diffusion is No Illusion - Neighbor Emulation in the Third Wave of Democracy." *Comparative Political Studies* 39(4):463–489.
- Buhaug, Halvard. 2005. "Dangerous Dyads Revisited: Democracies May Not Be That Peaceful After All." *Conflict Management and Peace Science* 22(2):95–111.
- Caprioli, Mary. 2005. "Primed for Violence: The Role of Gender Inequality in Predicting Internal Conflict." *International Studies Quarterly* 49(2):161–178.
- Caprioli, Mary and Peter F. Trumbore. 2005. "Rhetoric Versus Reality - Rogue States in Interstate Conflict." *Journal of Conflict Resolution* 49(5):770–791.

- Caprioli, Mary and Peter F. Trumbore. 2006. "Human Rights Rogues in Interstate Disputes, 1980–2001." *Journal of Peace Research* 43(2):131–148.
- Carpenter, Daniel P. and David E. Lewis. 2004. "Political Learning From Rare Events: Poisson Inference, Fiscal Constraints, and the Lifetime of Bureaus." *Political Analysis* 12(3):201–232.
- Cauthen, James N.G. and C. Scott Peters. 2003. "Courting Constituents: District Elections and Judicial Behavior on the Louisiana Supreme Court." *Justice System Journal* 24(3):265–281.
- Chamberlain, Robert and Donald P. Haider-Markel. 2005. "'Lien On Me': State Policy Innovation in Response to Paper Terrorism." *Political Research Quarterly* 58(3):449–460.
- Chang, Eric C.C. 2005. "Electoral Incentives for Political Corruption Under Open-List Proportional Representation." *Journal of Politics* 67(3):716–730.
- Chiozza, Giacomo. 2002. "Is There a Clash of Civilizations? Evidence From Patterns of International Conflict Involvement, 1946–1997." *Journal of Peace Research* 39(6):711–734.
- Choi, Seung-Whan and Patrick James. 2003. "No Professional Soldiers, No Militarized Interstate Disputes? A New Question for Neo-Kantianism." *Journal of Conflict Resolution* 47(6):796–816.
- Clark, David H. and Patrick M. Regan. 2003. "Opportunities to Fight - A Statistical Technique for Modeling Unobservable Phenomena." *Journal of Conflict Resolution* 47(1):94–115.
- Clark, David H. and Robert A. Hart. 1998. "Controlling Duration Dependence in Conflict Analyses: A Replication and Extension of "Regime Types and Status Quo Evaluations"." *American Journal of Political Science* 42(4):1335–1342.
- Colaresi, Michael and William R. Thompson. 2002. "Strategic Rivalries, Protracted Conflict, and Crisis Escalation." *Journal of Peace Research* 39(3):263–287.
- Crescenzi, Mark J.C. and Andrew J. Enterline. 2001. "Time Remembered: A Dynamic Model of Interstate Interaction." *International Studies Quarterly* 45(3):409–431.

- Crowley, Jocelyn Elise and Theda Skocpol. 2001. "The Rush to Organize: Explaining Associational Formation in the United States, 1860s–1920s." *American Journal of Political Science* 45(4):813–829.
- Davies, Graeme A.M. 2002. "Domestic Strife and the Initiation of International Conflicts - A Directed Dyad Analysis, 1950–1982." *Journal of Conflict Resolution* 46(5):672–692.
- Dickinson, Matthew J. and Kathryn Dunn Tenpas. 2002. "Explaining Increasing Turnover Rates Among Presidential Advisors, 1929–1997." *Journal of Politics* 64(2):434–448.
- Dixon, William J. and Paul D. Senese. 2002. "Democracy, Disputes, and Negotiated Settlements." *Journal of Conflict Resolution* 46(4):547–571.
- Dorussen, Han. 2006. "Heterogeneous Trade Interests and Conflict - What You Trade Matters." *Journal of Conflict Resolution* 50(1):87–107.
- Drury, A. Cooper. 2001. "Sanctions as Coercive Diplomacy: The U.S. President's Decision to Initiate Economic Sanctions." *Political Research Quarterly* 54(3):485–508.
- Edwards, Martin S. 2005. "Investor Responses to IMF Program Suspensions: Is Noncompliance Costly?" *Social Science Quarterly* 86(4):857–873.
- Enterline, Andrew J. and J. Michael Greig. 2005. "Beacons of Hope? The Impact of Imposed Democracy on Regional Peace, Democracy, and Prosperity." *Journal of Politics* 67(4):1075–1098.
- Firth, David. 1993. "Bias Reduction of Maximum Likelihood Estimates." *Biometrika* 80:27–38.
- Fox, John. 2000. *Nonparametric Simple Regression: Smoothing Scatterplots*. Thousand Oaks, CA: Sage.
- Fritz, Paul and Kevin Sweeney. 2004. "The (De)limitations of Balance of Power Theory." *International Interactions* 30(4):285–308.
- Gartzke, Eric and Quan Li. 2003a. "All's Well That Ends Well: A Reply to Oneal, Barbieri & Peters." *Journal of Peace Research* 40(6):727–732.
- Gartzke, Eric and Quan Li. 2003b. "Measure for Measure: Concept Operationalization and the Trade Interdependence-Conflict Debate." *Journal of Peace Research* 40(5):553–571.

- Gartzke, Eric and Quan Li. 2003c. "War, Peace, and the Invisible Hand: Positive Political Externalities of Economic Globalization." *International Studies Quarterly* 47(4):561–586.
- Gartzke, Eric, Quan Li and Charles R. Boehmer. 2001. "Investing in the Peace: Economic Interdependence and International Conflict." *International Organization* 55(2):391–438.
- Gelpi, Christopher and Joseph M. Grieco. 2001. "Attracting Trouble - Democracy, Leadership Tenure, and the Targeting of Militarized Challenges, 1919–1992." *Journal of Conflict Resolution* 45(6):794–817.
- 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):779–793.
- Goenner, Cullen F. 2004. "Uncertainty of the Liberal Peace." *Journal of Peace Research* 41(5):589–605.
- Goertz, Gary, Byran D. Jones and Paul F. Diehl. 2005. "Maintenance Processes in International Rivalries." *Journal of Conflict Resolution* 49(5):742–769.
- Goodcliffe, Jay and Darren G. Hawkins. 2006. "Explaining Commitment: States and the Convention Against Torture." *Journal of Politics* 68(2):358–371.
- Gowa, Joanne and Edward D. Mansfield. 2005. "Alliances, Imperfect Markets, and Major-Power Trade." *International Organization* 58(4):775–805.
- Hafner-Burton, Emilie M. and Alexander H. Montgomery. 2006. "International Organizations, Social Networks, and Conflict." *Journal of Conflict Resolution* 50(1):3–27.
- Heath, et al. 2005. "Women on the Sidelines: Women's Representation on Committees in Latin American Legislatures." *American Journal of Political Science* 49(2):420–436.
- Heinze, Georg and Michael Schemper. 2002. "A Solution to the Problem of Separation in Logistic Regression." *Statistics in Medicine* 21:2409–2419.
- Henderson, Errol A. and Richard Tucker. 2001. "Clear and Present Strangers: The Clash of Civilizations and International Conflict." *International Studies Quarterly* 45(2):317–338.

- Henisz, Witold J. 2002. "The Institutional Environment for Infrastructure Investment." *Industrial and Corporate Change* 11(2):355–389.
- Howard, Robert M. and Christine H. Roch. 2001. "Policy Change and the State Courts: The Case of Education Finance Reform." *Justice System Journal* 22(2):137–153.
- Humphries, Macartin. 2005. "Natural Resources, Conflict, and Conflict Resolution - Uncovering the Mechanisms." *Journal of Conflict Resolution* 49(4):508–537.
- James, Scott C. 2006. "Patronage Regimes and American Party Development From "The Age of Jackson" to the Progressive Era." *British Journal of Political Science* 36:39–60.
- Jungblut, Bernadette and Richard J. Stoll. 2002. "The Liberal Peace and Conflictive Interactions: The Onset of Militarized Interstate Disputes, 1950–1978." *Journal of Peace Research* 39(5):527–546.
- Ka, Sangjoon and Paul Teske. 2002. "Ideology and Professionalism - Electricity Regulation and Deregulation Over Time in the American States." *American Politics Research* 30(3):323–343.
- Kalbfleisch, John D. and Ross L. Prentice. 2002. *The Statistical Analysis of Failure Time Data, 2nd Edition*. Hoboken, NJ: Wiley-Interscience.
- Katz, Jonathan N. and Brian R. Sala. 1996. "Careerism, Committee Assignments, and the Electoral Connection." *American Political Science Review* 90(1):21–33.
- Keele, Luke J. 2008. *Semiparametric Regression for the Social Sciences*. New York, NY: John Wiley & Sons.
- Keller, Morton. 1977. *Affairs of State: Public Life in Late Nineteenth Century America*. Cambridge, MA: Harvard University Press.
- Kim, Hyung Min and David L. Rousseau. 2005. "The Classical Liberals Were Half Right (Or Half Wrong): New Tests of the "Liberal Peace"." *Journal of Peace Research* 42(5):523–543.
- Kinsella, David and Bruce Russett. 2002. "Conflict Emergence and Escalation in Interactive International Dyads." *Journal of Politics* 64(4):1045–1068.
- Krause, Volker. 2004. "Hazardous Weapons? Effects of Arms Transfers and Defense Pacts on Militarized Disputes, 1950–1995." *International Interactions* 30(4):349–371.

- Krutz, Glen S. 2005. "Issues and Institutions: "Winnowing" in the U.S. Congress." *American Journal of Political Science* 49(2):313–326.
- Lai, Brian. 2003a. "The Effects of Different Types of Military Mobilization on the Outcome of International Crises." *Journal of Conflict Resolution* 48(2):211–229.
- Lai, Brian. 2003b. "Examining the Goals of U.S. Foreign Assistance in the Post-Cold War Period, 1991–1996." *Journal of Peace Research* 40(1):103–128.
- Leblang, David A. 1999. "Domestic Political Institutions and Exchange Rate Commitments in the Developing World." *International Studies Quarterly* 43(4):599–620.
- Leblang, David and Steve Chan. 2003. "Explaining Wars Fought by Established Democracies: Do Institutional Constraints Matter?" *Political Research Quarterly* 56(4):385–400.
- Lebovic, James H. 2003. "The Limits of Reciprocity: Tolerance Thresholds in Superpower Conflict." *Journal of Peace Research* 40(2):139–158.
- Lebovic, James H. 2004. "Uniting for Peace? Democracies and United Nations Peace Operations After the Cold War." *Journal of Conflict Resolution* 48(6):910–936.
- Lektzian, David and Mark Souva. 2001. "Institutions and International Cooperation - An Event History Analysis of the Effects of Economic Sanctions." *Journal of Conflict Resolution* 45(1):61–79.
- Lemke, Douglas and William Reed. 1996. "Regime Types and Status Quo Evaluations: Power Transition Theory and the Democratic Peace." *International Interactions* 22:143–164.
- Lemke, Douglas and William Reed. 2001. "War and Rivalry Among Great Powers." *American Journal of Political Science* 45(2):457–469.
- Li, Quan and David Sacko. 2002. "The (Ir)relevance of Militarized Interstate Disputes for International Trade." *International Studies Quarterly* 46(1):11–43.
- Lujala, Päivi, Nils Petter Gleditsch and Elisabeth Gilmore. 2005. "A Diamond Curse? Civil War and a Lootable Resource." *Journal of Conflict Resolution* 49(4):538–562.
- Mansfield, Edward D. and Eric Reinhardt. 2003. "Multilateral Determinants of Regionalism: The Effects of GATT/WTO on the Formation of Preferential Trading Arrangements." *International Organization* 57(4):829–862.

- Mansfield, Edward D. and Jon C. Pevehouse. 2000. "Trade Blocs, Trade Flows, and International Conflict." *International Organization* 54(4):775–808.
- Mansfield, Edward and Jon Pevehouse. 2006. "Democratization and International Organizations." *International Organization* 60(1):137–167.
- Marinov, Nikolay. 2005. "Do Economic Sanctions Destabilize Country Leaders?" *American Journal of Political Science* 49(3):564–576.
- McDonald, Patrick J. 2004. "Peace Through Trade or Free Trade?" *Journal of Conflict Resolution* 48(4):547–572.
- McPherson, James M. 1988. *The Battle Cry of Freedom: The Civil War Era*. New York, NY: Oxford 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):180–193.
- Melander, Erik. 2005. "Gender Equality and Intrastate Armed Conflict." *International Studies Quarterly* 46(4):695–714.
- Meyer, Bruce D. 1990. "Unemployment Insurance and Unemployment Spells." *Econometrica* 58(4):757–782.
- Milner, Helen V. and K. Kubota. 2005. "Why the Move to Free Trade? Democracy and Trade Policy in the Developing Countries." *International Organization* 59(1):107–143.
- Mitchell, Sara McLaughlin and Brandon C. Prins. 2004. "Rivalry and Diversionary Uses of Forces." *Journal of Conflict Resolution* 48(6):937–961.
- Montinola, Gabriella R. 2003. "Who Recovers First? Banking Crises Resolution in Developing Countries." *Comparative Political Studies* 36(5):541–574.
- Mooney, Christopher Z. 2001. "Modeling Regional Effects on State Policy Diffusion." *Political Research Quarterly* 54(1):103–124.
- Mooney, Christopher Z. and Mei-Hsien Lee. 2000. "The Influence of Values on Consensus and Contentious Morality Policy: U.S. Death Penalty Reform, 1956–1982." *Journal of Politics* 62(1):223–239.

- Mousseau, Demet Yalcin. 2001. "Democratizing With Ethnic Divisions: A Source of Conflict?" *Journal of Peace Research* 38(5):547–567.
- Mousseau, Michael. 2000. "Market Prosperity, Democratic Consolidation, and Democratic Peace." *Journal of Conflict Resolution* 44(4):472–507.
- Mousseau, Michael. 2002. "An Economic Limitation to the Zone of Democratic Peace and Cooperation." *International Interactions* 28(2):137–164.
- Narendranathan, W. and M.B. Stewart. 1993. "How Does the Benefit Effect Vary as Unemployment Spells Lengthen." *Journal of Applied Econometrics* 8(4):361–381.
- Oneal, John R. 2003. "Measuring Interdependence and Its Pacific Effects: A Reply to Gartzke and Li." *Journal of Peace Research* 40(6):721–725.
- Oneal, John R. and Bruce M Russett. 1997. "The Classical Liberals Were Right: Democracy, Interdependence, and Conflict, 1950–1985." *International Studies Quarterly* 41(2):267–293.
- Oneal, John R. and Bruce Russett. 1999a. "Assessing the Liberal Peace With Alternative Specifications: Trade Still Reduces Conflict." *Journal of Peace Research* 36(4):423–442.
- Oneal, John R. and Bruce Russett. 1999b. "Is the Liberal Peace Just an Artefact of Cold War Interests? Assessing Recent Critiques." *International Interactions* 25(3):213–241.
- Oneal, John R. and Bruce Russett. 2005. "Rule of Three, Let It Be? When More Really Is Better." *Conflict Management and Peace Science* 22(4):293–310.
- Palmer, Glenn, Tamar R. London and Patrick M. Regan. 2004. "What's Stopping You?: The Sources of Political Constraints on International Conflict Behavior in Parliamentary Democracies." *International Interactions* 30(1):1–24.
- Palmer, Harvey D. and Guy D. Whitten. 2000. "Government Competence, Economic Performance and Endogenous Election Dates." *Electoral Studies* 19(2–3):413–426.
- Peceny, Mark, Caroline Beer and Shannon Sanchez-Terry. 2002. "Dictatorial Peace?" *American Political Science Review* 96(1):15–26.
- Peet, Curtis and Mark V. Simon. 2000. "Social Purpose and Militarized Disputes." *Journal of Peace Research* 37(6):659–677.

- Pickering, Jeffrey. 2002. "Give Me Shelter: Reexamining Military Intervention and the Monadic Democratic Peace." *International Interactions* 28(4):293–324.
- Powers, Kathy L. 2004. "Regional Trade Agreements as Military Alliances." *International Interactions* 30(4):373–395.
- Prentice, R.L. and L.A. Gloeckler. 1978. "Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data." *Biometrics* 34(1):57–67.
- Rasler, Karen A. and William R. Thompson. 2001. "Rivalries and the Democratic Peace in the Major Power Subsystem." *Journal of Peace Research* 38(6):659–683.
- Rasler, Karen A. and William R. Thompson. 2004. "The Democratic Peace and a Sequential, Reciprocal, Causal Arrow Hypothesis." *Comparative Political Studies* 37(8):879–908.
- Rasler, Karen A. and William R. Thompson. 2006. "Contested Territory, Strategic Rivalries, and Conflict Escalation." *International Studies Quarterly* 50(1):145–167.
- Ravlo, Hilde, Nils Petter Gleditsch and Han Dorussen. 2003. "Colonial War and the Democratic Peace." *Journal of Conflict Resolution* 47(4):520–548.
- Ray, James Lee. 2003. "Explaining Interstate Conflict and War: What Should Be Controlled For?" *Conflict Management and Peace Science* 20(2):1–31.
- Reed, William. 2000. "A Unified Statistical Model of Conflict Onset and Escalation." *American Journal of Political Science* 44(1):84–93.
- Reed, William. 2003a. "Information and Economic Interdependence." *Journal of Conflict Resolution* 47(1):54–71.
- Reed, William. 2003b. "Information, Power, and War." *American Political Science Review* 97(4):633–641.
- Ruppert, David, M.P. Wand and R.J. Carroll. 2003. *Semiparametric Regression*. Cambridge, UK: Cambridge University Press.
- Santner, Thomas J. and Diane E. Duffy. 1986. "A Note on A. Albert and J.A. Anderson's Conditions for the Existence of Maximum Likelihood Estimates in Logistic Regression Models." *Biometrika* 73(3):755–758.

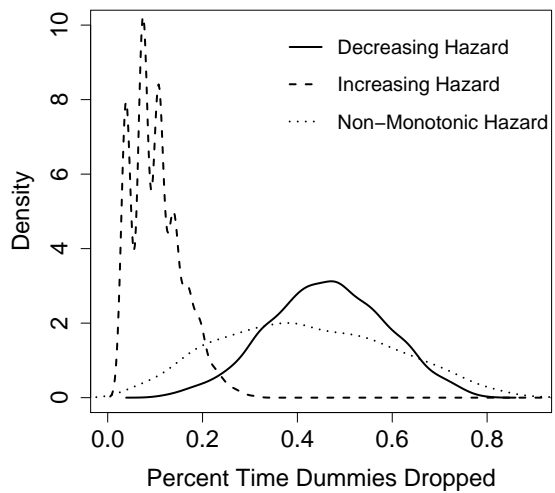
- Schamis, Hector E. and Christopher Way. 2003. "Political Cycles and Exchange Rate-Based Stabilization." *World Politics* 56(1):43–.
- Schlesinger, Arthur M. 1944. "Biography of a Nation of Joiners." *American Historical Review* 50:1–25.
- Schultz, Kenneth A. 2001. *Democracy and Coercive Diplomacy*. Cambridge, UK: Cambridge University Press.
- Searle, Shayle R. 1982. *Matrix Algebra Useful for Statistics*. New York, NY: John Wiley & Sons.
- Sechser, Todd S. 2004. "Are Soldiers Less War-Prone Than Statesmen?" *Journal of Conflict Resolution* 48(5):746–774.
- Senese, Paul D. 2005. "Territory, Contiguity, and International Conflict: Assessing a New Joint Explanation." *American Journal of Political Science* 49(4):769–779.
- Senese, Paul D. and John A. Vasquez. 2003. "A Unified Explanation of Territorial Conflict: Testing the Impact of Sampling Bias, 1919–1992." *International Studies Quarterly* 47(2):275–298.
- Sherman, Richard. 2001. "Democracy and Trade Conflict." *International Interactions* 27(1):1–28.
- Simmons, Beth A. 2000. "International Law and State Behavior: Commitment and Compliance in International Monetary Affairs." *American Political Science Review* 94(4):819–835.
- Skocpol, Theda, Marshall Ganz and Ziad Munson. 2000. "A Nation of Organizers: The Institutional Origins of Civic Voluntarism in the United States." *American Political Science Review* 94:527–546.
- So, Ying. 1995. "A Tutorial on Logistic Regression." *SUGI Proceedings* .
- Sobek, David. 2005. "Machiavelli's Legacy: Domestic Politics and International Conflict." *International Studies Quarterly* 49(2):179–204.
- Sørli, Mirjam E., Nils Petter Gleditsch and Håvard Strand. 2005. "Why Is There So Much Conflict in the Middle East?" *Journal of Conflict Resolution* 49(1):141–165.

- Stalley, Phillip. 2003. "Environmental Scarcity and International Conflict." *Conflict Management and Peace Science* 20(2):33–58.
- Stein, Jana Von. 2005. "Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance." *American Political Science Review* 99(4):611–622.
- Sweeney, Kevin J. 2003. "The Severity of Interstate Disputes - Are Dyadic Capability Preponderances Really More Pacific?" *Journal of Conflict Resolution* 47(6):728–750.
- Sweeney, Kevin and Paul Fritz. 2003. "Jumping on the Bandwagon: An Interest-Based Explanation for Great Power Alliances." *Journal of Politics* 66(2):428–449.
- Thacker, Strom Cronan. 1999. "The High Politics of IMF Lending." *World Politics* 52(1):38–.
- Volden, Craig. 2002. "The Politics of Competitive Federalism: A Race to the Bottom In Welfare Benefits?" *American Journal of Political Science* 46(2):352–363.
- Volden, Craig. 2006. "States as Policy Laboratories: Emulating Success in the Children's Health Insurance Program." *American Journal of Political Science* 50(2):294–312.
- Volden, Craig and Clifford J. Carruba. 2004. "The Formation of Oversized Coalitions in Parliamentary Democracies." *American Journal of Political Science* 48(3):521–537.
- Walter, Barbara F. 2004. "Does Conflict Beget Conflict? Explaining Recurring Civil War." *Journal of Peace Research* 41(3):371–388.
- Wiebe, Robert H. 1967. *The Search for Order, 1877–1920*. New York, NY: Hill and Wang.
- Yatchew, Adonis and Zvi Griliches. 1985. "Specification Error in Probit Models." *The Review of Economics and Statistics* 67(1):134–139.
- Zorn, Christopher. 2001. "Estimating Between and Within Cluster Covariate Effects, With an Application to Models of International Disputes." *International Interactions* 27(4):433–445.
- Zorn, Christopher. 2005. "A Solution to Separation in Binary Response Models." *Political Analysis* 13:157–170.

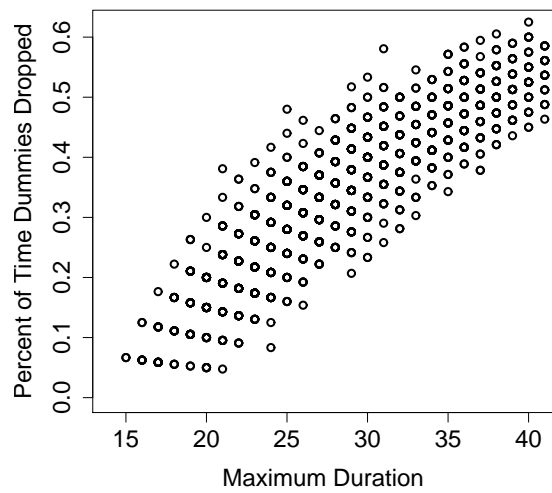
Table 3: Crowley and Skocpol Logit Replication

	Time Dummies	$t, t^2, t^3$	Auto Spline
Constant	<b>-2.44</b> (.43)	<b>-2.19</b> (.43)	<b>-2.09</b> (.42)
Urban Growth	.11 (.06)	.10 (.06)	.09 (.06)
Manufacturing Per Capita	-.06 (.29)	-.15 (.29)	-.17 (.29)
Railroad Mi. Per Capita	-18.83 (10.64)	-18.48 (10.52)	-19.01 (10.58)
Teachers Per Capita	-9.49 (9.53)	-7.13 (8.68)	-7.42 (8.71)
Percent Literate	.055 (.053)	.040 (.053)	.036 (.053)
Percent in Union Armies	<b>.029</b> (.010)	<b>.031</b> (.010)	<b>.031</b> (.010)
Pension \$ Per Pensioner	<b>.48</b> (.08)	<b>.38</b> (.08)	<b>.39</b> (.07)
Electoral Competitiveness	<b>.045</b> (.016)	<b>.043</b> (.016)	<b>.043</b> (.016)
Foreign Born Growth	.028 (.09)	.004 (.08)	.002 (.08)
Population Growth	-.0017 (.0011)	-.001 (.001)	-.001 (.001)
Odd Fellows Per Capita	-3.66 (6.69)	.35 (6.56)	.92 (6.55)
Percent Protestant	.056 (.029)	<b>.072</b> (.028)	<b>.075</b> (.028)
Neighbor Effects	<b>.32</b> (.05)	<b>.28</b> (.04)	<b>.28</b> (.04)
$t$		<b>.55</b> (.18)	
$t^2$		<b>-.23</b> (.09)	
$t^3$		.014 (.014)	
Log-Likelihood	-1508.48	-1544.44	-1546.59
N =	2529	2529	2529

Standard errors in parentheses . Bold estimates:  $p < .05$

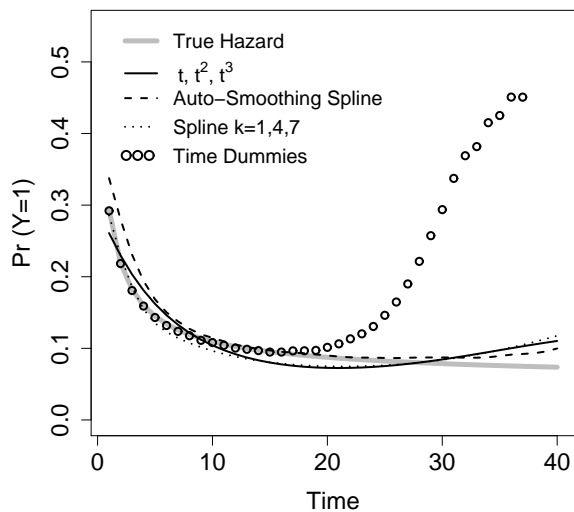


(a) By Type of Hazard

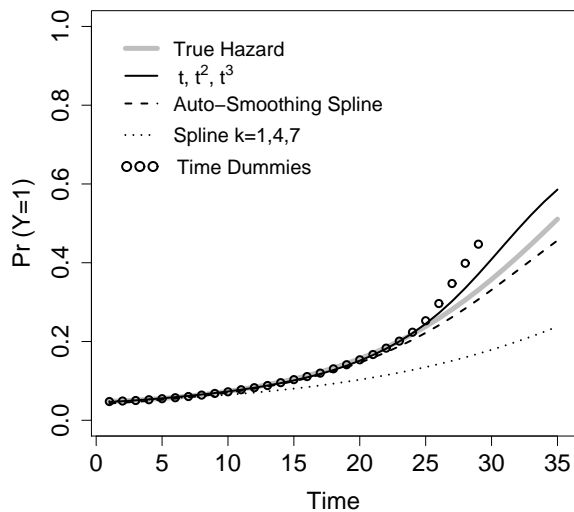


(b) As a Function of Maximum Duration

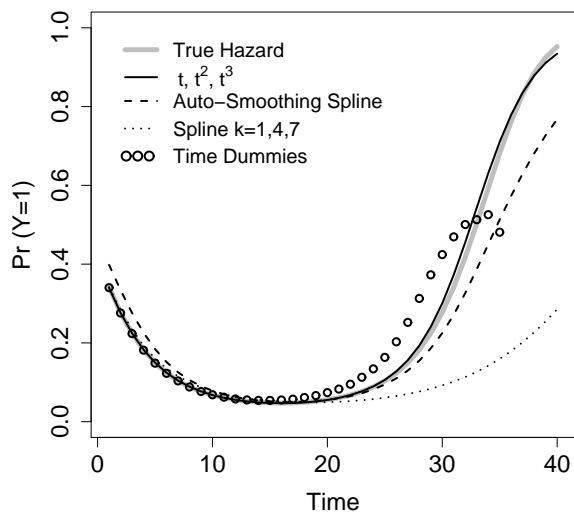
Figure 2: Percent of Time Dummies Dropped Due to Quasi-Complete Separation



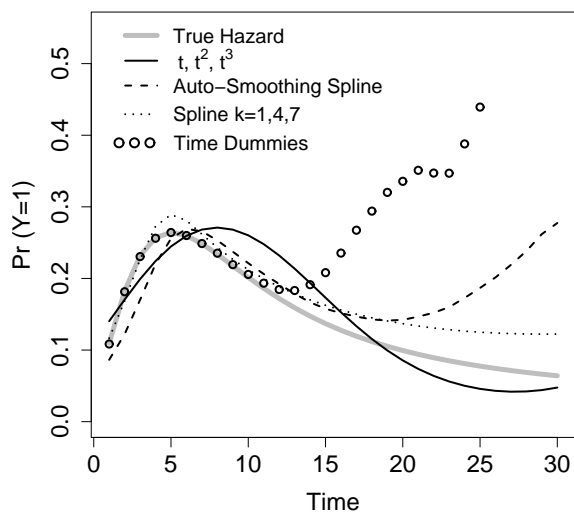
(a) Decreasing Hazard



(b) Increasing Hazard



(c) Non-Monotonic Hazard 1



(d) Non-Monotonic Hazard 2

Figure 3: Monte Carlo Comparison

Figure 4: Effect of Time (Since Civil War) on Associational Formation

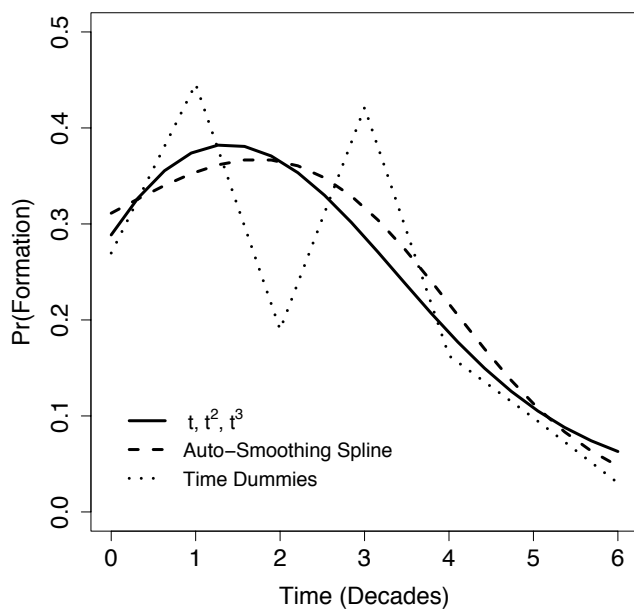
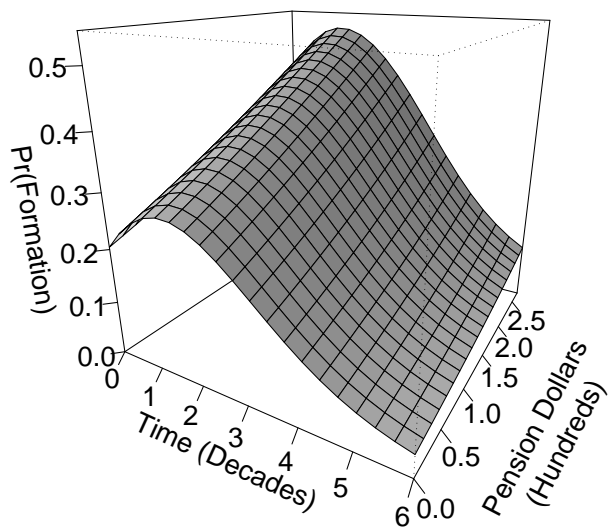
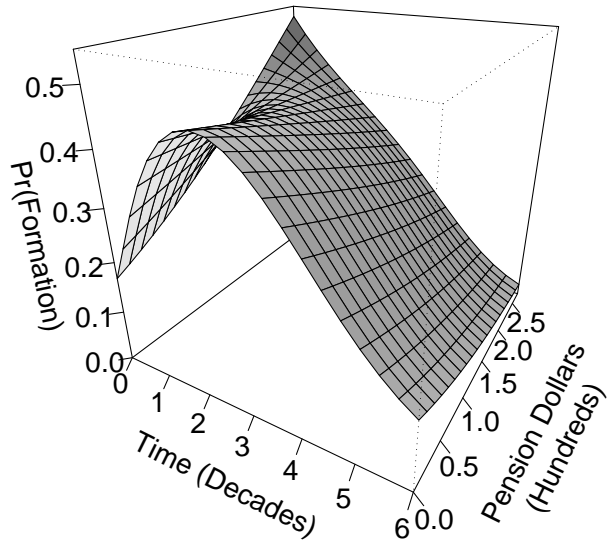


Figure 5: Effect of Pension Dollars and Time ( $t$ ,  $t^2$ ,  $t^3$  Model)



(a) Proportional Hazard



(b) Nonproportional Hazard