Advance Access publication June 15, 2010

Political Analysis (2010) 18:271–292
doi:10.1093/pan/mpq013

Back to the Future: Modeling Time Dependence in Binary Data

David B. Carter
Department of Political Science, Pond Laboratory 211, The Pennsylvania State University, University Park, PA 16802
e-mail: dbc10@psu.edu (corresponding author)

Curtis S. Signorino
303 Harkness Hall, Department of Political Science, University of Rochester, Rochester, NY 14627,
e-mail: curt.signorino@rochester.edu

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 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 temporal dependence. 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 autosmoothing 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 the historical-institutionalist perspective.

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

Authors' note: We thank Daehee Bak, Hein Goemans, Jay Goodliffe, Luke Keele, Eduardo Leoni, Arthur Spirling, Randy Stone, Chris Zorn, and three anonymous reviewers for their comments and suggestions. This article has also benefited from comments provided during seminars at the University of Berkeley, the University of Minnesota, the University of Pittsburgh, and the Watson Center for Conflict and Cooperation, as well as from comments provided during conference presentations at the annual meetings of the Political Methodology Society (2007), Peace Science Society (2007), and the American Political Science Association (2008). We thank Jocelyn Crowley, Theda Skocpol, and Glenn Palmer for generously providing their data. Supplementary materials (Web Appendix) for this article are available on the *Political Analysis* Web site.

Table 1 Use of splines and time dummies

	No hazard (%)	*Interpret hazard*	
Splines	88 (96.7)	3 (3.3)	91
Time dummies	24 (85.7)	4 (14.3)	28
	112 (94.1)	7 (5.9)	119

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 that scholars should ''take time seriously.'' 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 temporal dependence in discussions of empirical results.

Consider Table 1, which summarizes all published articles we found that follow BKT's advice on using time dummies or splines.[1] 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 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 temporal dependence 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 one of many available spline techniques—procedures for creating a ''smooth'' relationship between two variables (here, between the dependent variable and time). However, most substantive researchers do not seem to understand what splines are. This is suggested in part by the fact that many researchers employ

[1]We compiled this list by locating all published articles that cited BKT in the Social Sciences Citation Index 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.

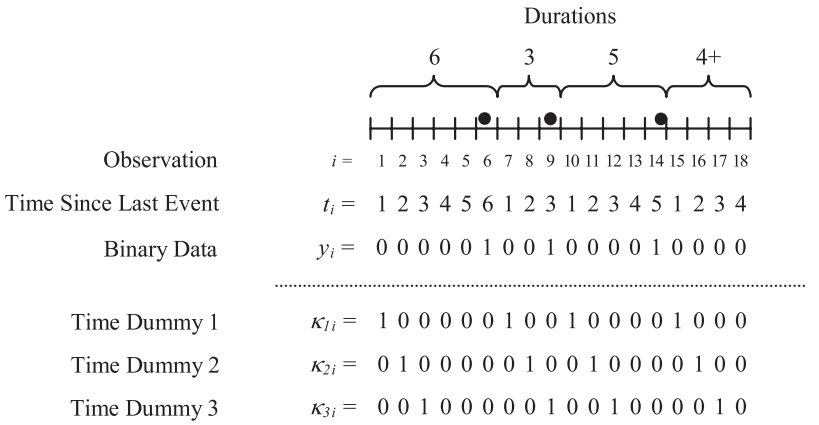

Fig. 1 Binary representation of duration data.

the parameters (e.g., "knot" values) that BKT use, regardless of whether they are appropriate for the researchers' data. And, as Table 1 shows, the dependent variable's temporal dependence, controlling for the regressors, is rarely ever discussed when splines are employed.

In this article, 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.[2] As we later show, the cubic polynomial approximation does not cause the same data problems as time dummies. Moreover, the cubic polynomial is related to splined time but much easier to implement and to interpret. Indeed, modeling nonproportional hazards is relatively straightforward with the cubic polynomial approach. In the online Web Appendix to this article, we provide R and Stata code for both the cubic polynomial and spline approaches.

This article 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. Next, 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 its analysis using a nonproportional hazards model. Our nonproportional hazards version provides new, richer empirical support for the theory of Crowley and Skocpol (2001).

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 refers 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 the temporal component 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

[2]Here, "t" refers to the time (or duration) since the last event was observed (e.g., war).

last right-censored observation of length 4. Vast literatures exist on duration (or survival) analysis (see, e.g., 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 nonmonotonic with time. Alternatively, researchers sometimes opt for the (semi-parametric) Cox model.

Now consider the binary data shown below the time line and denoted as y_i. As BKT notes (and Alt, King, and Signorino [2001] demonstrates), 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 shows, derivation of the binary data version of a Cox model is relatively straightforward (Prentice and Gloeckler 1978; Katz and Sala 1996; Beck, Katz, and Tucker 1998; Meyer 1990; Narendranathan and Stewart 1993; Cameron and Trivedi 2005, 601–3). Let us assume the longest duration is T periods. For observation i, let x_i be a $1 \times k$ row vector of k covariate values, t index the time since the last event (i.e., duration so far), and β be a $k \times 1$ vector of parameters. If one starts with the assumption of a Cox relative risk model with hazard

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

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

$$\Pr(y_i = 1|x_i, \kappa_i) = 1-\exp(-\exp(x_i\beta+\kappa_i\alpha)), \tag{2}$$

where κ_i is a $1 \times T$ row vector of time dummy values $[\kappa_{1i}\, \kappa_{2i} \ldots \kappa_{Ti}]$, and α is a $T \times 1$ vector of coefficients associated with the time dummies.

It is important to note that the ''time dummies'' are not just the time variable t. Rather, time dummies are duration-specific fixed effects, where each time dummy represents a particular duration $t \in \{1, 2, 3, \ldots, T\}$. Returning to Figure 1, we have displayed three transposed time dummies below the dotted line. For example, κ_1 will be one whenever $t = 1$ and 0 otherwise. Similarly, κ_2 will equal one whenever $t = 2$ and 0 otherwise. In general, one needs a κ_τ for each observed value of 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 ten years, then there will be ten time dummies. If it is thirty, then there will be thirty time dummies. The researcher can include all time dummies and drop the constant or include the constant and drop 1 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 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 nonmonotonic). In principle, this allows consistent β estimates, even if the true shape of the hazard is unknown (Meyer 1990). As BKT demonstrates, the logit model very closely approximates the cloglog when

the probability that $y = 1$ is relatively low. However, if the percentage of observations for which $y = 1$ is relatively high, cloglog and logit can diverge significantly.[3] One particularly attractive aspect of cloglog for many researchers is the fact that it is consistent with the popular Cox proportional hazards model. However, BKT correctly points out that there is usually no good reason to privilege the cloglog distribution over logit. Therefore, BKT recommends that researchers use logit (or probit) since they are so widely available. This is not really an issue anymore since cloglog is now widely available in statistical packages. Nevertheless, we will maintain the use of logit throughout this article.

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. A maximum duration of 30 years in the data will often be represented by 30 time dummy parameters. Assuming the hazard is relatively smooth, the time dummies approach is inefficient compared to splines or the cubic polynomial approach—the latter requiring estimation of only 3 additional parameters. Of course, as with any estimation problem, 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. As BKT also notes, inefficiency problems 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 β, such that for all observations $x_i\beta > 0$ when $y_i = 1$ and $x_i\beta < 0$ 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 β 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).[4]

These three cases, complete separation, quasi-complete separation, and overlap, are mutually exclusive. However, only in the case of overlap do noninfinite maximum likelihood estimates (MLEs) exist (Albert and Anderson 1984). If the data exhibit either complete or quasi-complete separation, no MLEs exist and MLE routines will push the value of the estimated $\hat{\beta}$ for the offending regressors to ∞ 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

[3] In the online Web Appendix, we provide Monte Carlo experiments that assess logit's performance when the data-generating process is cloglog.

[4] 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 nontechnical explanation. See Zorn (2005) for examples of overlap versus separation with a continuous regressor.

Table 2 Separation in binary data with three time dummies

Observation i	t	κ_1	κ_2	κ_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

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: κ_1, κ_2, and κ_3. As the time index t shows, the data consist 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.

Consider the first time dummy, κ_1. We see from Table 2 that there is no value of κ_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, κ_2, reveals the same type of overlap between values of κ_2 and y. Now consider κ_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 κ_3 perfectly predicts $y = 1$. Before proceeding with the regression, it would drop κ_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 κ_3 to be pushed toward ∞, 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 regard, we might ask: (1) Under what conditions are 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 κ_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,

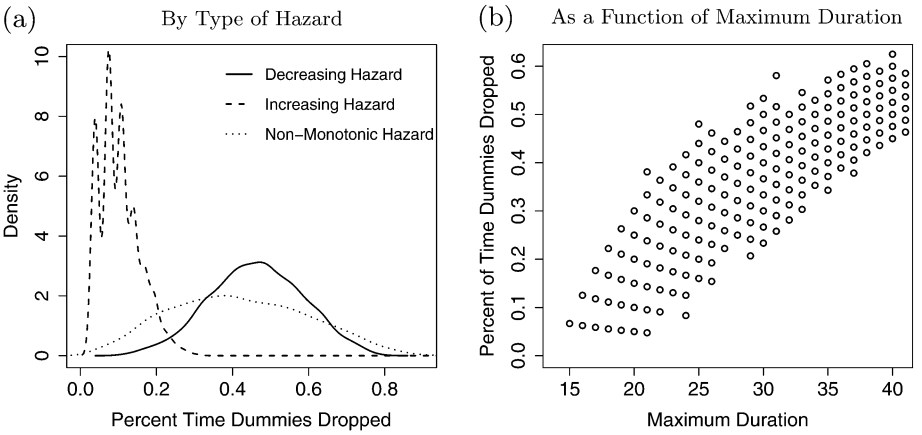

Fig. 2 Percent of time dummies dropped due to quasi-complete separation.

splines, cubic polynomial). The Monte Carlo experiments were conducted using various reasonable examples of decreasing, increasing, and nonmonotonic hazards.[5] For the time being, consider Figure 2, which displays the percent of time dummies needed to be dropped due to separation in the Monte Carlo analyses. Figure 2a shows the density of the percent of time dummies dropped under each hazard assumption. In both the decreasing and nonmonotonic hazard scenarios, up to 80% of the time dummies are dropped, with averages of around 45% and 35% dropped, respectively. The increasing hazard case is not nearly as problematic. Yet, even there, Figure 2a shows that up to 25% of time dummies were dropped in some Monte Carlo simulations.

Figure 2b provides some insight into the results in Figure 2a. Each circle in Figure 2b displays the percent of time dummies dropped in a particular Monte Carlo experiment as a function of the maximum duration in the data.[6] As the graph shows, there is a fairly straightforward relationship: data sets 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. Nonmonotonic hazards may also suffer from similar problems but that will depend on the form of nonmonotonicity (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, data sets where n is not

[5]The nonmonotonic hazard here is referred to as ''nonmonotonic hazard 2'' in Section 4. This hazard has a sharp rise in failure followed by a gradual decrease over time.

[6]Figure 2b is based on the nonmonotonic hazard. The graphs for the increasing hazard and decreasing hazard scenarios look similar and are available in the Web Appendix.

overly large (i.e., $n < 10{,}000$) 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 the data by Oneal and Russett in BKT, where 3 of 34 time dummies are dropped, along with 916 observations. A very large n of 20,990 explains why the percentage dropped, ≈ 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 the models. For instance, in model 5 of Lemke and Reed (1996), 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. One potential solution is not to include the dummies that induce separation but to "smooth" across these missing dummies when producing a hazard plot. This approach would produce a similar hazard plot in the replication of the study by Oneal and Russett. However, it would be much less appropriate in any study that resembles Palmer, London, and Regan. Alternatively, researchers can also aggregate adjacent durations to avoid separation. Thus, in the example provided in Table 2, a researcher would estimate κ_1 and κ_{23}, where κ_{23} is the sum of κ_2 and κ_3. Although either of these suggestions could be helpful, both seem unnecessary given that neither a cubic polynomial nor a spline suffers from separation issues. Additionally, we demonstrate below that time dummies also do a poor job relative to a cubic polynomial or spline in estimating the hazard. This finding is consistent with the point made by Beck, Katz, and Tucker (1998, 1270) that time dummies perform poorly due to imprecise κ estimates unless N is very large.

3 Love the Spline, Hate the Spliner

The second approach to modeling time dependence that BKT advocates is to use splines. Like Cox models, time dummies can produce a "bumpy" hazard plot. BKT, therefore, recommends the use of natural cubic splines when researchers want a relatively smooth plot of the hazard. 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 in statistical packages such as Stata, SPSS, and 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. 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). Because of the complexity of splines, 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 article. We instead refer readers to Keele (2008) for a recent introduction to splines by a political scientist. However, some intuition concerning splines will be helpful for at least two reasons. First, as Table 1 indicates, existing work on splines does not provide enough detail for users to comfortably 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 Web Appendix) R and Stata programs 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 procedure 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 spline techniques is that they generate a set of vectors that are a function 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.[7] 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 and some smooth function of time $s(t_i)$. We can specify the logit equation for this situation as

$$\Pr(y_i = 1|x_i, t) = \frac{1}{1+\exp[-(x_i\beta+s(t_i)).]}.$$

[7]Suppose we have some vector space S such that it is possible to express every vector in S as a linear combination of k $(x_1, x_2, \ldots, x_k)$ 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.

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

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

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

The simple cubic spline is conceptually very similar:

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

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 β and α coefficients via logistic regression.

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 β and α 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 provides 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). We want to interpret the temporal dependence in our dependent variable, conditional on the regressors. More roughly put, we want to interpret the effect of time t on the estimated $\Pr(y_i = 1 | x_i, t)$. Of course, time is not an

independent entity that acts on our dependent variable. In this case, it is a stand-in (like a dummy variable) for something we have not yet modeled functionally or via the included regressors.

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 $\tilde{t} = \{1, 2, 3, \ldots, \max(t)\}$. 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 $\Pr(y_i = 1|x_i, \tilde{t})$ 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 (nonspline) 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 to plot the estimated $\Pr(y_i = 1|x_i, \tilde{t})$ with $\tilde{t}$ along the x axis. In the Web Appendix, we provide R and Stata programs that demonstrate 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 η_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 versus t and chooses knot locations where the relationship appears to change. Ruppert, Wand, and Carroll (2003, 57–72) illustrate at length how a researcher can choose knot locations based upon a visual examination of two-dimensional scatterplots. 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, $\Pr(y_i = 1|x_i, t)$, 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 theory and/or 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 nonmonotonic 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, 270–1), 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 (analysis of Oneal and Russett [1997] 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 $t = 1, 4$, and 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 overfitting. Although a more detailed discussion of automated smoothing techniques is beyond the scope of this article, we have included examples in the plots of the Monte Carlo analyses in Section 4.[8] 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.[9] To make this concrete, suppose a researcher with regressors x_i wanted to conduct logistic regression, control for temporal dependence, and interpret the effect of time on $\Pr(y_i = 1|x_i, t)$. Using this approach, her logit equation would take the form

$$Pr(y_i = 1|x_i, t) = \frac{1}{1+\exp\left[-\left(x_i\beta+\alpha_1 t_i+\alpha_2 t_i^2+\alpha_3 t_i^3\right)\right]}, \tag{5}$$

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 semiparametric models such as the Cox proportional hazard model. At the same time, the cubic polynomial avoids the overfitting 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 temporal dependence 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 $\Pr(y_i = 1|x_i, t)$, rather than interpreting the individual polynomial terms. Having estimated the model, the researcher simply creates a new time

[8] 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.

[9] See also Efron (1988) for an application of a cubic polynomial in estimating Kaplan-Meier survival curves using binomial count data.

vector $\tilde{t} = \{1, 2, 3, \ldots, \max(t)\}$, along with its square and cube, $\tilde{t}^2$ and $\tilde{t}^3$, respectively. The fitted values for $\Pr(y_i = 1|x_i, \tilde{t})$ 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 $\Pr(y_i = 1|x_i, \tilde{t})$ 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 demean t before squaring and cubing it. This will reduce (although not completely eliminate) collinearity. Second, 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 15,625. The solution here is no different than what common practice should already be. Simply examine the range of all the variables in the data, as well as t, t^2, and t^3, and then rescale variables as necessary by dividing by some multiple of 10. We have found that using $\frac{t}{100}$ and its square and cube generally works quite well. Similarly, t, t^2, and $\frac{t^3}{1000}$ also work well. In either case, users should remember to use the appropriate scaling when plotting the estimated $Pr(y_i = 1|x_i, \tilde{t})$.

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. This allows us to approximate very closely an underlying Cox framework. Moreover, time dummies 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

$$Pr(y_i = 1|x_i, \kappa_i) = \frac{1}{1+\exp(-(x_i\beta+\kappa_i\alpha))}, \tag{6}$$

where for observation i, x_i consists of a constant and a single regressor value that is uniformly drawn between -2 and 2. All β's are set equal to one. As in equation (2), κ_i is a $1 \times T$ row vector of time dummy values for observation i. The shape of the hazard is therefore determined by the α parameters associated with the time dummies.

We conduct Monte Carlo experiments for constant, decreasing, increasing, and two different nonmonotonic hazards. In each, we run 10,000 Monte Carlo iterations with samples of $n = 2000$. 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, and 7 for the B-spline estimation.[10]

Because BKT already addresses the effect of ignoring temporal dependence on the estimates associated with substantive regressors, we do not focus on that here. Moreover, our Monte Carlo analyses confirm an interesting result shown in one of Box-Steffensmeier and Jones (2004, 91) empirical replications: that the modeling choice for time dependence has little effect on the substantive regressors' parameter estimates (i.e., the β's above), so long as one implements a ''reasonable'' modeling technique (e.g., B-splines, automatic smoothing via GCV, a cubic polynomial, or time dummies when separation is not extreme). Choice of modeling technique has more effect on our estimation and interpretation of the hazard itself.

Figure 3 demonstrates 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

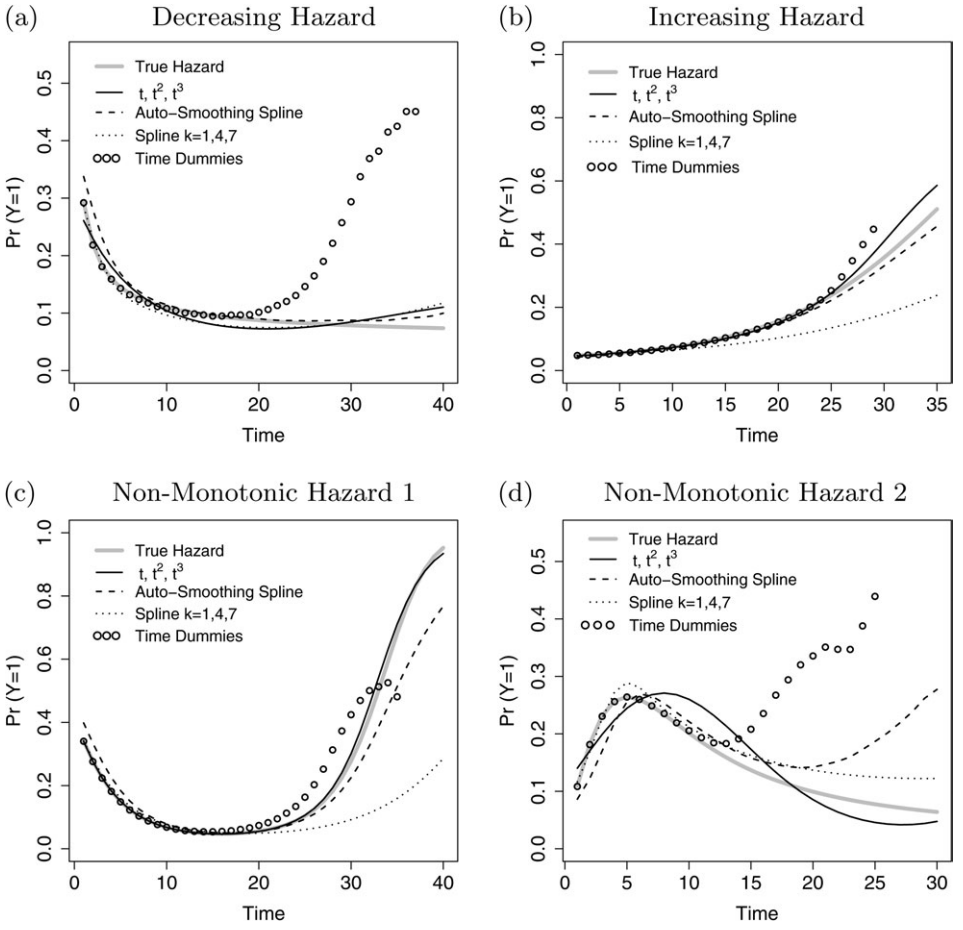

Fig. 3 Monte Carlo comparison.

[10]A number of variations of the above were analyzed. For example, in one set of analyses, we assumed the data generating process was cloglog—consistent with a Cox proportional hazard framework. Those results were virtually identical to those presented here. We also ran Monte Carlos to assess the impact of a time-trended covariate. Again, those results were very similar to those reported here. All these are provided in our Web Appendix.

line, the cubic polynomial approximation by a solid black line, the automated smoothing spline by a dashed line, the B-spline (with knots at $t = 1, 4, 7$) by a dotted line, and the time dummies model by the open circles.

Consider first Figure 3a, 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 3a the effect of those separation problems. Most problematic, we would on average infer that a nonmonotonic hazard existed since the time dummies plot begins increasing sharply at about $t = 20$. Second, 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.[11]

Figure 3b displays the results when the true hazard is increasing in time. In this case, the cubic polynomial and the autosmoothing spline both perform very well, with the autosmoothing 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 3b. 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, and 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 3c and 3d show the results for two different nonmonotonic hazards. The u-shaped hazard in Figure 3c 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. Mortality in Western countries takes a form similar to this hazard. As in the two preceding cases, the cubic polynomial performs very well. The autosmoothing spline fares reasonably well, as do the time dummies. For the time dummies, the increasing hazard later in time prevents the really problematic separation issues seen in Figure 3a. However, notice the (somewhat disturbing) fact that time dummies produce a u-shaped hazard in both Figure 3a (where it is incorrect) and 3(c) (where it is correct). Finally, the B-spline with knots at $t = 1$, 4, and 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 nonmonotonic hazard in Figure 3d 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

[11]Confidence intervals have been calculated for all graphs in Figure 3. These are available in the Web Appendix.

$t = 12$. 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 $t = 1, 4$, and 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 autosmoothing 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 the analysis of Crowley and Skocpol (2001) of why, when, and how membership associations (such as the Young Men's Christian Association, or YMCA) formed in the United States.[12] Crowley and Skocpol (2001) focuses on two competing perspectives. The first perspective, popular among historians, is that socioeconomic modernization and high levels of immigration were the main catalysts of associational development (Wiebe 1967; Berthoff 1971; Keller 1977). In contrast, Crowley and 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 aftereffects of association building and development during the U.S. Civil War (see Crowley and Skocpol [2001, 814–6]). 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) presents a new data set that tracks the establishment of state-level chapters for 21 organizations in the 48 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

[12] We have replicated a number of other analyses: for example, the reanalysis of Oneal and Russett (1997) by Beck, Katz, and Tucker (1998), as well as the analysis in Palmer, London, and Regan (2004). In all cases, we find that a cubic polynomial performs just as well as splines with automatic knot placement. Results are available in our Web Appendix.

Table 3 Crowley and Skocpol logit replication

	Time dummies	*t, t^2, t^3*	*Auto spline*
Constant	**2.44** (0.43)	**2.19** (0.43)	**2.09** (0.42)
Urban growth	0.11 (0.06)	0.10 (0.06)	0.09 (0.06)
Manufacturing per capita	−0.06 (0.29)	−0.15 (0.29)	−0.17 (0.29)
Railroad mile 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	0.055 (0.053)	0.040 (0.053)	0.036 (0.053)
Percent in union armies	**0.029** (0.010)	**0.031** (0.010)	**0.031** (0.010)
Pension $ per pensioner	**0.48** (0.08)	**0.38** (0.08)	**0.39** (0.07)
Electoral competitiveness	**0.045** (0.016)	**0.043** (0.016)	**0.043** (0.016)
Foreign born growth	0.028 (0.09)	0.004 (0.08)	0.002 (0.08)
Population growth	−0.0017 (0.0011)	−0.001 (0.001)	−0.001 (0.001)
Odd fellows per capita	−3.66 (6.69)	0.35 (6.56)	0.92 (6.55)
Percent protestant	0.056 (0.029)	**0.072** (0.028)	**0.075** (0.028)
Neighbor effects	**0.32** (0.05)	**0.28** (0.04)	**0.28** (0.04)
t		**0.55** (0.18)	
t^2		**−0.23** (0.09)	
t^3		0.014 (0.014)	
Log-likelihood	−1508.48	−1544.44	−1546.59
N	2529	2529	2529

Note. SEs in parentheses. Bold estimates: $p < .05$.

(2001) estimates 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. As Table 3 shows, the substantive results 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.[13]

6.1 *Associational Formation over Time*

Although 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 socioeconomic 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

[13] 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, whereas we 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.

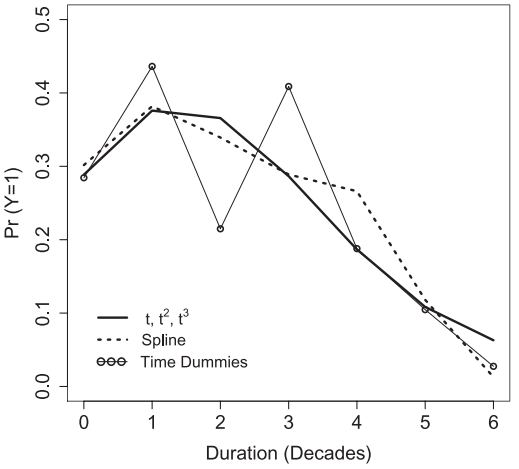

Fig. 4 Effect of time (since civil war) on associational formation.

the socioeconomic 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 30 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, code for the average pension received by Union Civil War pensioners in a given state. As Crowley and Skocpol (2001, 818) notes, 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.

Figure 5a 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 5a is based on a proportional hazards model. As Box-Steffensmeier and Zorn (2001) notes, nearly all duration models commonly used by political scientists assume proportional hazards.[14]

[14]For a technical discussion of the differences between proportional and nonproportional hazard models, see Kalbfleisch and Prentice (2002), Box-Steffensmeier and Jones (2004, 131–7), or Box-Steffensmeier and Zorn (2001).

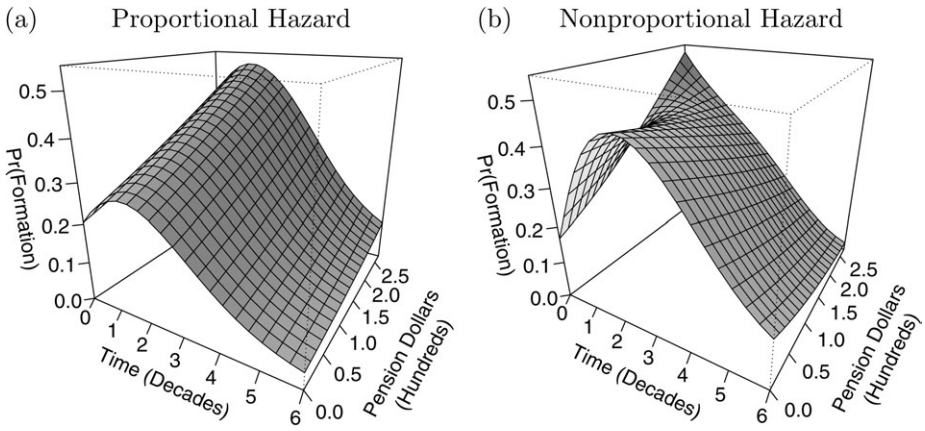

Fig. 5 Effect of pension dollars and time (t, t^2, t^3 model).

Indeed, all the models in Table 3 are proportional hazard models (technically, proportional odds models). When the proportional hazards assumption is incorrect, our parameter estimates and SEs may be substantially biased. Moreover, there are many instances where our theory suggests the hazards are *nonproportional*.

Fortunately, modeling nonproportional hazards is relatively easy when using a cubic polynomial or splines. To model nonproportional hazards with a cubic polynomial, we recommend simply interacting the regressors of interest with t, t^2, and t^3 (or appropriately scaled versions thereof). 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.[15]

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 to 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 pension dollars variable with each of the t, t^2, and t^3 variables. Although not shown, the parameter estimates on the main regressors remain similar to those in the restricted model.[16] A likelihood ratio test rejects the restrictions at well below the 0.05 level. Thus, the nonproportional hazard model appears warranted in this case.

[15]Technically, a researcher is testing the joint null hypothesis of proportionality for a regressor and that the hazard is approximated by a cubic polynomial.
[16]Estimates for the unrestricted model are reported in the Web Appendix.

The graph in Figure 5b 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 5a, 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 5b 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 5a (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 5b. 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 70 years, most have already formed chapters. The lowest pension states lag behind, so there are still chapters to be formed.

In sum, although Crowley and Skocpol is correct about the positive impact of Union armies, it does 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 dependence. However, most substantive researchers do not appear to understand splines.

Evidence for this supposition is the fact that most researchers use BKT's application-specific knots of 1, 4, and 7, whether or not this parameter selection is appropriate for the researcher's own application. More often than not, researchers omit any analysis of the temporal dependence in their data. Researchers who want to ''taking time seriously'' should do more than just add time regressors as controls. Researchers should also plot and interpret the hazard. In this article, we have tried to clarify how one might do that following BKT's recommendation to use splines.

Although we strongly support the recommendation to use splines, we have also suggested a related, but simpler alternative: using a cubic polynomial approximation (t, t^2, and t^3). This technique is easy to implement and to interpret. It can also model a wide variety of hazard shapes. Our Monte Carlo analysis (and numerous empirical replications) suggests that 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 the work by Crowley and Skocpol (2001) on federated voluntary associational formation in the postbellum United States. In fact, our more nuanced analysis provides even more support for their theory.

Finally, although this article tries to assist researchers in modeling time dependence in binary data, it does so in only a very basic way. Interpreting the hazard can offer clues as to what processes or regressors might be omitted from the existing model. The replication of Crowley and Skocpol (2001) is instructive in this sense. The larger point here is that researchers need to put more work, theoretical and empirical, into properly modeling and interpreting temporal dynamics. Indeed, for many areas of substantive research, it may be that none of the models here is even approximately ''correct.'' If, for instance, structural breaks occur over time or the process has an autoregressive structure, then different modeling techniques will be needed. Thus, even though we argue that logit with a smooth function of time will work in *many* scenarios, we encourage researchers to explore more sophisticated models of temporal dynamics (e.g., dynamic choice models) when theory suggests it is appropriate.

Funding

National Science Foundation (SES-0213771 and SES-0413381).

References

Albert, A., and J. A. Anderson. 1984. On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* 71:1–10.

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* 9:1–24.

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* 42:1260–88.

Berthoff, Rowland. 1971. *An unsettled people: Social order and disorder in American history.* New York: Harper and Row.

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* 45:972–88.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and applications*. Cambridge, UK: Cambridge University Press.

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* 42:1335–42.

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* 45:813–29.

Dorussen, Han. 2006. Heterogeneous trade interests and conflict—What you trade matters. *Journal of Conflict Resolution* 50:87–107.

Efron, Bradley. 1988. Logistic regression, survival analysis, and the Kaplan-Meier curve. *Journal of the American Statistical Association* 83:414–25.

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.

Goertz, Gary, Byran D. Jones, and Paul F. Diehl. 2005. Maintenance processes in international rivalries. *Journal of Conflict Resolution* 49:742–69.

Heinze, Georg, and Michael Schemper. 2002. A solution to the problem of separation in logistic regression. *Statistics in Medicine* 21:2409–19.

Kalbfleisch, John D., and Ross L. Prentice. 2002. *The statistical analysis of failure time data*. 2nd ed. Hoboken, NJ: Wiley-Interscience.

Katz, Jonathan N., and Brian R. Sala. 1996. Careerism, committee assignments, and the electoral connection. *American Political Science Review* 90:21–33.

Keele, Luke J. 2008. *Semiparametric regression for the social sciences*. New York: John Wiley & Sons.

Keller, Morton. 1977. *Affairs of state: Public life in late nineteenth century America*. Cambridge, MA: Harvard University Press.

Lemke, Douglas, and William Reed. 1996. Regime types and status quo evaluations: Power transition theory and the democratic peace. *International Interactions* 22:143–64.

McPherson, James M. 1988. *The battle cry of freedom: The civil war era*. New York: Oxford University Press.

Meyer, Bruce D. 1990. Unemployment insurance and unemployment spells. *Econometrica* 58:757–82.

Narendranathan, W., and M. B. Stewart. 1993. How does the benefit effect vary as unemployment spells lengthen. *Journal of Applied Econometrics* 8:361–81.

Oneal, John R., and Bruce M. Russett. 1997. The classical liberals were right: Democracy, interdependence, and conflict, 1950–1985. *International Studies Quarterly* 41:267–93.

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* 30:1–24.

Prentice, R. L., and L. A. Gloeckler. 1978. Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* 34:57–67.

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* 73:755–8.

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: John Wiley & Sons.

Senese, Paul D. 2005. Territory, contiguity, and international conflict: Assessing a new joint explanation. *American Journal of Political Science* 49:769–79.

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* 94:527–46.

So, Ying. 1995. *A tutorial on logistic regression.* Cary, NC: SAS White Papers.

Wiebe, Robert H. 1967. *The search for order, 1877–1920*. New York: Hill and Wang.

Yatchew, Adonis, and Zvi Griliches. 1985. Specification error in probit models. *The Review of Economics and Statistics* 67:134–9.

Zorn, Christopher. 2005. A solution to separation in binary response models. *Political Analysis* 13:157–70.