Long-Term Effects in Models with Temporal Dependence

Appendix

Laron K. Williams
University of Missouri
williamslaro@missouri.edu

Overview

In this document I provide more in-depth support for a series of claims that I make in the manuscript related to the relationship between functional form and long-term effects (LTE) with a variety of estimation techniques, Monte Carlo experiments examining the performance of estimators (with and without non-proportional hazards), the effects of compression on substantive effects, and additional analyses of Clare (2010).

Illustration of Long-Term Effects with Hazard Rates of Various Shapes

To demonstrate how the shape of long-term effects is related to the underlying hazard rate, I establish four scenarios of underlying hazard rates. For guidance, I follow the approach taken by Carter and Signorino (2010) and establish four scenarios based on the relationship with *time*: increasing, decreasing, and two scenarios of non-monotonic relationships.

I first simulate four data sets of 1000 observations, with the binary dependent variable based on an explanatory variable (x) drawn from a uniform distribution, $\in [-2, 2]$ ($\beta = 1$) and a hazard rate with the logistic link function:

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

I depict these four hazard rates in Figure S.1. The positive and negative hazards are derived from a Weibull distribution, whereas the two non-monotonic hazards are derived from a parabolic and log-logistic distribution, respectively.

[Figure S.1 about here]

Recall that there are two obvious approaches (Hanmer and Kalkan 2012) to calculating the long-term effects, given the possibility of hyper-conditionality. The first of these is to set up multiple simulation scenarios to demonstrate the variability of substantive effects at different baseline probabilities. The second is to calculate long-term effects for all observed values and then average those values to produce an "average" long-term effect. Both of these techniques are shown in the manuscript.

The first functional form is negative and is probably the most common pattern of temporal dependence, simply because models of civil wars and international conflicts typically follow these patterns. In Figure S.2 I show four long-term effects calculated at different scenarios of t. For the four scenarios, I change the value of *time* in the following formula:

$$\text{LTE}_{\mathbf{X_C}}^{t+1} = \text{Pr}(\hat{y} = 1|\mathbf{X_{C2}}, time = 0) - \text{Pr}(\hat{y} = 1|\mathbf{X_C 1}, time = t_{X_C}) \tag{2}$$

For example, in the first scenario ($t_{X_C} = 1$) the long-term effect at time $t + 1$ is the probability immediately following the event occurring at time t (*time* = 0) compared to the probability for the scenario plus 1 (1). At time $t + 2$, I compare the probability at time 1 to 2, and so on. The second scenario ($t_{X_C} = 5$) presents a larger long-term effect at time $t+1$ because it is the probability when *time* = 0 compared to *time* = 5. The third and fourth scenarios have even larger long-term effects at time $t + 1$ because they compare the probability at *time* = 0 to *time* = 10 and 20, respectively. The ranges of the x-axes of these latter figures are smaller so that we do not extrapolate beyond the maximum value of time.

[Figures S.2-S.5 about here]

The first inference from the case with negative duration dependence (FigureS.2) is that the shape of the long-term effect is related to the underlying hazard rate. Since the hazard rate is linearly decreasing (up to about $t = 15$), then the long-term effect for those time periods will be positive. The second inference is that the magnitude of the long-term effects depends on the simulation scenario, and more specifically, the value of *time* in the simulation scenario. In the case of negative duration dependence, higher values of *time* in the simulation scenario ($t_{\mathbf{X_C}}$) will produce more positive LTEs because the hazard rate when *time* = 0 is much larger than the hazard rate at higher values of *time*. We can also infer that the long-term effect of the event has a lasting effect that elevates the risk of another event for nearly 15 years.

In Figure S.3 we depict the same scenarios for the case of positive duration dependence. Since the underlying hazard rate rises with time (Figure S.1), occurrence of the event at time t decreases the risk of another event. This decrease is statistically significant (for the $t_{X_C} = 1$ scenario) after four time periods and the negative LTEs grow stronger with time.

Figures S.4 and S.5 depict LTEs for non-linear functional forms. Even though the functional forms are more complex than the simple ones derived from the Weibull model, the figures are interpreted in a similar manner. Moreover, the interpretation of the long-term effects is simple if one considers that the sign, shape and magnitude of the LTEs are the result of subtracting the probability at a particular value of *time* from the probability when *time* = 0. For example, in the non-monotonic 1 functional form, $t_{X_C} = 1$ scenario, the underlying hazard rate is highest at time *time* = 1, so the long-term effect will be positive until the hazard becomes flat.

[Figures S.6-S.9 about here]

Figures S.10-S.13 compare two versions of depicting LTE with different functional forms. While the right panels appear similar, they differ in one notable way. They provide greater context for the LTE because they list the value of t in the baseline scenario. The left panels explicitly provide the two values that make up the moving differences depicted in the right panels: Pr$(\hat{y} = 1|y_t = 1, \mathbf{X_C})$ and Pr$(\hat{y} = 1|y_t = 0, \mathbf{X_C})$. The vertical confidence intervals are slightly jittered on the x-axis to allow for easy comparison of the scenarios. The first non-monotonic scenario (Figure S.8) depicts a particularly interesting LTE that fluctuates from being negative to not statistically different from zero and then positive.

[Figures S.10-S.13 about here]

Monte Carlo Experiments

Long-term effects—much like other quantities of interest—"assume that the statistical model is identified and correctly specified" (King, Tomz and Wittenberg 2000: 351). More specifically, correct estimation of LTE depends on both the estimate of the coefficient of interest (β) and the hazard rate ($\beta_{\mathbf{t}}$). However, it is not yet clear how inaccurately specifying the temporal dependence influences our inferences regarding those coefficients and the calculation of LTE.

Scholars face a multitude of choices when specifying the variables in the hopes of modeling temporal dependence, ranging from temporal dummy variables to cubic polynomials to a variety of splines. I will show that this choice has minimal consequences for the estimation of LTE; on the whole, as long as one's inferences are limited to the common values of t and the hazard rate is estimated in a reasonable fashion, there are few differences in LTEs across these estimation techniques. In this section I will describe the Monte Carlo experiments and highlight a number of important findings that illuminate the use of LTE in practice.

My goal is to illustrate the performance of a variety estimation techniques to model temporal dependence under different conditions. I generate a variety of data sets ranging in size ($N \in$

$\{1000, 5000, 10000\}$) with the binary dependent variable based on an explanatory variable (X_K) drawn from a uniform distribution from -2 to 2 ($\beta_{X_K} = 1$), a negative constant ($\beta_C \in \{-3, -2\}$, to get a variety of frequencies of events) and a hazard rate with the logistic link function:[1]

$$\Pr(y_{i,t} = 1|\mathbf{x_{i,t}}) = \frac{1}{1 + \mathrm{e}^{-(\mathbf{x_{i,t}}\beta + h(t))}} \tag{3}$$

I create four functional forms of hazard rates based on their relationship with time: increasing, decreasing, and two scenarios of non-monotonic relationships (see Figure S.1).

We can assess the performance of these functional forms under different hazard rates with two simple criteria that directly inform the calculation of LTE: bias in the coefficient of the variable of interest (β_{X_K}), and by comparing the estimates to the true hazard rate at different values of t.

[Tables S.1-S.4 about here]

In Tables S.1-S.4 I show the performance of five estimation techniques: exponential (flat, or no hazard), time dummies (Beck, Katz and Tucker 1998), cubic polynomials (Carter and Signorino 2010), B-splines located at three knots (1, 4, and 7, to compare to the default from many statistical programs), and automated smoothing splines (via generalized cross-validation). In the first two columns I show the absolute bias and mean squared error of $\beta_{X_K} = 1$. In the final two columns I assess whether the standard errors accurately portray the variability of the estimate by comparing the mean standard error of β_{X_K} to the standard deviation of the 1000 estimates of β_{X_K}. Better estimators of β_{X_K} are those that have lower values of bias and where the average standard errors are closer to the standard deviations of the estimated β_{X_K} (Carsey and Hardin 2014: 84-96).

The evidence is clear; failing to correctly model temporal dependence—in a variety of functional forms—biases the coefficient for the explanatory variable of interest.[2] In nearly every single scenario under different functional forms, the exponential distribution has the highest absolute bias and mean squared error. While the coefficients are biased downward, the similarity of the average standard error and the standard deviation of the coefficients means that the standard errors—in these circumstances[3]—reflect the true variability of the estimates.

Morever, as long as one specifies the temporal dependence in a reasonable fashion (thus excluding the exponential functional form, which assumes constant hazards), then one is likely to retrieve the actual estimate of β_{X_K}, on average. Carter and Signorino (2010: 284; see also Box-Steffensmeier and Jones 2004: 91) stress this point as well, but it is reassuring that the inference from the key explanatory variable will be accurate as long as some care is taken in model specification. There is little difference between the estimation techniques in terms of bias and standard errors, so one will have to look elsewhere to adjudicate between the techniques.

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

[2] This is consistent with omitted variable bias present in probit models that ignore temporal dependence (Yatchew and Griliches 1985).

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

Of course, this is not to say that the LTE will be accurate under all circumstances, as this depends on appropriately modeling the underlying hazard rate as well. By comparing the true hazard rate to the estimates from the various techniques, we can assess whether the estimates contain large amounts of bias.

[Figures S.14-S.17 about here]

Figures S.14-S.17 provide these comparisons for four different data sets based on sample size and size of the constant. The average hazard rates are quite close to the true hazard rate over the entire period for the scenarios where the constant is -3 (Figure S.14 and S.16). Furthermore, the estimation techniques do a solid job of picking up non-linearities in the hazard when they occur. The estimates reflect the slight increase at higher values in the first non-monotonic scenario and the curvilinear shape at the lower values of the second non-monotonic scenario.

More generally, the estimates are closer to the true hazard rate at lower values of t. As t increases, the estimates stray further from the true hazard and this is exacerbated in two situations. First, when the constant increases (from -3 to -2) there is a higher percentage of 1s in the sample (from about 15% to about 40%), at which point there is a greater divergence from the true hazard at lower values. Second, some techniques provide better estimates than others. Cubic polynomials do a particularly poor job at higher values of t and they often depict non-linearities in the hazard rate that are not there (as shown in the *Non-Monotonic 2* scenario in Figure S.15).[4]

This is easy to see in Figures S.18-S.21 where I depict the true hazard rate compared to the 1000 estimates in the scenario where all the techniques performed most admirably (i.e., N=1000, β_C=-3).

[Figures S.19-S.21 about here]

At low values of t the hazard rates have much less variation on average than at higher rates. One can therefore be much more confident that an estimate at lower, more common values of t will be more efficient.

In practice, however, there is still reason to be optimistic about the use of these techniques to address temporal dependence. First, BTSCS models with temporal dependence often explain rather rare events such as civil wars or interstate conflicts, and these data sets are more similar to the scenarios depicted in Figures S.14 and S.16 than S.15 and S.17. In situations where there is a lower percentage of 1s in the sample, scholars should have more confidence that a reasonable technique will provide a close estimate of the true hazard rate. Second, in the simulated data sets, values of t that are larger than about 20 are relatively rare (all of the data are right-skewed), so it is reasonable to expect that our estimates of the hazard rate under more anomalous situations would be potentially wrong. If scholars are cautious about drawing inferences from observations that are not representative of the entire sample, then they will focus their inferences on low to moderate values of t. In this range of values the estimates are quite close to the truth.

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

Long-Term Effects and Non-Proportional Hazards

A very real possibility when dealing with BTSCS models is that the effects of an explanatory variable may fluctuate as a function of time. In event history analysis, much has been written about the consequences of violating the proportional hazards assumption on model estimation (Box-Steffensmeier and Jones 2004) and interpretation (Licht 2011). In the BKT approach, models that only include the time variables (e.g., t, t^2 and t^3) assume that the effects of X_K on Y do not vary meaningfully across time. Examples of research questions where we might unknowingly violate this assumption abound, especially in international relations. One could imagine that the effects of an unresolved territorial dispute would have the largest effect immediately following the previous conflict, and that this effect would shrink with time. Appropriate modeling of non-proportional hazards would entail interacting the offending variable with the measures of time. It is not clear, however, how the presence of non-proportional hazards complicates the calculation and depiction of long-term effects. In the manuscript I presented an example of how to depict the long-term effects, and this section addresses the remaining component—how violating the NPH assumption affects the calculation of LTE.

I repeat the data-generating process outlined in the above section for the case with 1,000 simulations and a constant of -3, which causes about 16-25$\%$ of the outcomes to be coded as 1. In addition to having similiar features as most international conflict data sets, this scenario represents favorable conditions to assessing the effects of NPH since the estimation techniques performed well (see above). The crucial exception is that the influence of X_K now varies as a function of t ($X_K \times t$), and the influence varies across scenarios: ($\beta_{Xt} \in \{0.2, 0.1, 0.04, 0.02\}$). In the case of negative non-proportional hazards these coefficients are negative.

Tables S.5 and S.6 detail the performance of the β_{X_K} and its standard error across the various strengths of non-proportional hazards and estimation techniques.

[Tables S.5-S.6 about here]

There is little difference in the performance across estimation techniques (though the automated smoothing splines appear to be slight favorites). The bias in β_{X_K} tends to be highest when the non-proportional hazards are strongest (i.e., at $\beta_{Xt} = -0.2$) for the negative non-proportional hazards case (Table S.6) but not for the positive case (Table S.5). Overall, it looks as though if one estimates an additive model in the face of non-proportional hazards there will be some bias in β_{X_K}, but it does not look consequential since the average β_{X_K} is still quite close to 1. It remains to be seen whether the calculation of the hazard rate and the effects of X_K across time will be as forgiving.

Figure S.22 provides the average hazard rates of the four estimation techniques from the additive (i.e., proportional hazards) model compared to the true hazard rate (with a slight degree of non-proportionality, β_{Xt}=0.4 and -0.4, respectively).

[Figure S.22 about here]

As expected, the slight non-proportionality is not enough to make the additive estimates diverge substantially from the true hazard. This is especially the case at low values of t.

Of course, the results from Tables S.5-S.6 and Figure S.22 provide only a snapshot of how the BKT techniques perform in the face of non-proportional hazards. This is because Tables S.5-S.6 and Figure S.22 illustrate performance of the estimates of β_{X_K} and $h(t)$ under the unique circumstances when $t = 0$ and $X_K = 0$, respectively (Licht 2011). Scholars are not usually concerned with making inferences about such a small subset of cases. Instead, scholars are interested in the inferences regarding how the effects of X_K change as a function of t. This inference is more clearly illustrated by examining the average predictive differences (APD)—the average change in the probability that $\hat{y} = 1$ given an increase in X_K from 0 to 1—across values of t.

[Figures S.23-S.24 about here]

Figure S.23 (S.24) shows the APD across values of t for four estimation techniques under negative (positive) duration dependence with varying levels of positive (negative) non-proportional hazards. Keep in mind that these APD are estimated from models that ignore the non-proportional hazards that are present in the data. Thus, it is admirable that the additive models capture some of the changing influence of X_K over time. In Figure S.23 this is evident by the slight increase in the APD at the higher values of t. This is even more obvious in Figure S.24 where there is enough variation to have estimates across a longer range of t. All four estimation techniques pick up on the negative non-proportional hazards which eventually decrease the magnitude of the APD at moderate values of t.

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

Government Fractionalization and Hyper-Conditionality

Based on the statistical significance of the coefficient for *ideological distance*, Clare finds support for his "overall theoretical premise that ideological outlier parties can use their credible threats to defect from the coalition, and the increased bargaining leverage these threats provide, to shift a government's foreign behavior toward their desired policy positions" (2010: 981). To further demonstrate the variation in conflict behavior within coalition governments, Clare compares the probability of dispute initiation of a "fractionalized coalition" (i.e., one whose values of *ideological* *distance* vary) to a "cohesive coalition" and a single-party majority government. I recreate these figures in Figure S.25, and include a rug plot of the distribution of *ideological distance* within the sample.

[Figure S.25 about here]

He points out that "fractionalized coalitions with left-wing outlier parties are less likely to initiate disputes than either cohesive coalitions (panel a) or single-party governments (panel b). Alternatively, those with a far right-wing outlier party are more likely to initiate conflicts than their cohesive coalitional or single-party counterparts" (Clare 2010: 981). These two probabilities are statistically different (at the 90% confidence level) at values of *ideological distance* less than -108 and greater than 92 for the left panel, and less than -115 for the right panel. Since there are only 14 observations in the data set (Greece 1989 and Denmark 1957-1960) that have these values, I would caution scholars from making inferences about these relatively rare values.

Substantive Effects and Scenario Selection

The degree of hyper-conditionality present in the two panels in Figure 4 in the manuscript provides a warning against selecting only one scenario to depict substantive effects.[5] It is important to note that the effects of *ideological distance* vary much more across the values of *peaceyears* (i.e., along the x-axis) than across observations with the same value of *peaceyears*. Thus, the magnitude of the first difference depends much more on the values of the temporal dependence variables than the geostrategic control variables such as military capabilities, alliances or contiguity.[6] It is therefore evident that the practice of interpreting the results based on one scenario can produce widely varying predicted effects.

The statistical significance of changes in key theoretical variables will also vary based on the values of the simulation scenarios. Ai and Norton (2003) show that the first difference of a variable may only be statistically significant for a subsection of the observations. The next step in exploring hyper-conditionality is to determine whether the statistically significant observations were concentrated in certain ranges of values for *peaceyears*. Even though the coefficient for *ideological distance* is statistically significant at the 99% confidence level, due to compression, only about 22% of the observations have values of Y^* that are close enough to the center of the cumulative density function to actually produce a statistically significant difference. Indeed, all of the observations where the change in *ideological distance* is statistically significant (at the 95% confidence level)

[5] In his project, Clare (2010: 981) interprets the effects of *ideological distance* on the probability of dispute initiation for a scenario where all the control variables (including *peaceyears*) are held at their means.

[6] To get a sense of the variation in the first differences that is solely explained by temporal dependence, I regressed first differences on the three *peaceyears* variables (shown in the *Appendix* file). The adjusted R^2 is 0.84, which supports my contention that the vast majority of variation in the first difference is attributable to the values of the temporal dependence variables. Furthermore, a heteroskedastic regression model points to the error variance decreasing as a function of *peaceyears* (β=-0.03, p-value < 0.001), which suggests that there is a great deal more variation in the possible effects of *ideological distance* when conflict is in the recent past.

occur within seven years since the previous dispute. This by itself is not particularly surprising, as it places unreasonable demands on the theory that the effects are statistically significant in each and every case. At the same time, this sort of exploration might lead to inferential improvements by encouraging a reanalysis of the conditions under which these effects reach conventional levels of statistical significance.

Given the nonlinear nature of probit models, and the substantial effects of temporal dependence variables, it is instructive to examine whether the differences between scenarios are more or less pronounced for different conflict histories. This will help determine whether different types of coalitions consistently behave differently, or if these differences are merely a product of the scenarios chosen to calculate the effects. In Figure 6 in the manuscript I purposely identify four scenarios where the baseline probability varies considerably to demonstrate the influence of compression on substantive effects. The values of the variables depicted in those scenarios are provided in Table S.7.

[Table S.7 about here]

Model Fit

The model of dispute initiation that Clare (2010) provides fits the data quite well, with an expected percent correctly predicted (ePCP) of 94.6%. Figure S.26 demonstrates how well the model performs in terms of correctly predicting disputes and non-disputes across different probability thresholds (King and Zeng 2006). The area under the curve (AUC) is 0.77, with a 95% confidence interval of [0.71, 0.82] (Weidmann and Ward 2010).

[Figures S.26-S.27 about here]

Figure S.27 provides the separation plot based on the predicted probabilities and observed outcomes (Greenhill, Ward and Sacks 2011).

Non-Proportional Hazards in Clare (2010)

In the manuscript I claim that the substantive effects of *government fractionalization* on *dispute initiation* are highly dependent on the values of the temporal dependence variables in the simulation scenarios. In order to claim that these effects are due to the methodological artifact of compression, I need to eliminate the possibility that it is actually due to non-proportional hazards. If the effects of *government fractionalization* vary across *time since previous dispute* due to an explicit interactive relationship and I omit the interactions, then I am likely to observe substantial variability.

To ensure that the variability of the substantive effects are caused by compression rather than non-proportional hazards, I interact the *peaceyears* cubic polynomials with *government fractionalization*. I present the results in Table S.8.

[Table S.8 about here]

None of the interactions are statistically significant, which suggests that the effects of *government fractionalization* do not vary as a function of *time since previous dispute*. More importantly, I fail to reject the null hypothesis that the three interactions are jointly equal to 0 at conventional levels of statistical significance using both the Wald test ($\chi^2_3 = 2.2$, p-value=0.53) and likelihood ratio test ($\chi^2_3 = 2.6$, p-value=0.46). Thus, it is clear that the variability in the substantive effects of *government fractionalization* in Clare (2010) is due to compression rather than unmodeled non-proportional hazards.

Survey of Temporal Dependence Variables in Top Journals

I conducted a survey of all the articles published in the *American Political Science Review* (through February 2014), *American Journal of Political Science* (through July 2014), *Journal of Politics* (through April 2014) and *International Studies Quarterly* (through June 2014) that cited Beck, Katz and Tucker (1998). For each article, I noted the type of temporal dependence variable (i.e., dummy variables, splines, polynomials, etc), whether they were statistically significant, the manner of interpretation (if any), the quantities of interest of key variables (and the simulation scenarios used), and whether they discussed and interpreted long-term effects.

Out of 178 articles, 131 employ some version of temporal dependence variables. Of these, only two discuss the possibility of lasting effects of the independent variables (Bennett 2006; Beardsley 2008), but neither calculates the actual long-term effects. Furthermore, even though compression changes the size of the quantities of interest, only 4.7% (or 6 out of 129) show how the calculated quantities of interest vary across multiple simulation scenarios.

Temporal Dependence Variables and the CDF

Temporal dependence variables are nearly ubiquitous in BTSCS models of political phenomena in the last two decades (see the review in Carter and Signorino 2010). While they are typically deployed as a methodological fix to the statistical nuisance of duration dependence, I argue that there is the unintended consequence of hyper-conditionality. Of course, in a logit model there is always varying substantive effects due to compression, so what makes the problems associated with temporal dependence variables more severe?

I argue that the long ranges of the *time* variable, coupled with the substantively meaningful coefficients, result in a wide range of possible values along the CDF. This means that the same observation can experience widely different effects of a theoretical variable because of the values of the temporal dependence variables in the simulation scenario. To illustrate this result, I take the four functional forms of the duration dependence described above (Figure S.1) and generate $X\beta$ values "purged" of the effects of the temporal dependence variables. I then select one observation where the baseline probability of the dependent variable is close to 0.5. Then, simply by varying the values of the temporal dependence variables, we can explore how they determine the observation's location along the CDF and ultimately the size of the quantities of interest.

[Figure S.28 about here]

Figure S.28 demonstrates that, across a variety of functional forms of duration dependence, an observation's location along the CDF is dependent on the values of the temporal dependence variables. Ultimately, the magnitude of the changes in predicted probabilities that one infers from these models depends on the particular simulation scenarios.

Variability of Substantive Effects

In the manuscript I make two claims regarding the variability of substantive effects for BTSCS models with varying ranges of probabilities. The first claim is that the average size of the change in predicted probabilities will be largest at probabilities closer to 0.5. The second claim is that the variability of the substantive effects depends on the range of predicted probabilities.

The change in predicted probabilities for a 1-unit increase in some variable, x_k, is a function of both the size of the effect, β_{x_k}, and the location of the observation along the CDF (which determines the probability that $y = 1$). The formula for this change is the following (Agresti 1996: 104):

$$\Delta \text{Pr}(y = 1|\Delta x_k) = \Delta x_k \beta_{x_k} \text{Pr}(y = 1)(1 - \text{Pr}(y = 1)) \tag{4}$$

To demonstrate how the range of proabilities influence the size of these changes, I set up a series of simulations of 2000 observations drawn from a uniform distribution of probabilities. Thus, the sample size and standard deviations are nearly constant across simulations. I then calculate the changes with Equation 4 for a $\beta_{x_k} = 1$. I present the distributions of the substantive effects for the various simulations as box-whisker plots in Figures S.29 and S.30.

[Figures S.29-S.30 about here]

Figure S.29 illustrates these two claims about variability. First, the average change in predicted probabilities of a 1-unit shift in x, given a $\beta_{x_k} = 1$, is much higher for the range of probabilities from 0.4 to 0.6. As the range of probabilities shifts farther away from 0.5, the average magnitude decreases. Second, the variability of these changes is much smaller in the middle range of probabilities ([0.4, 0.6]) compared to those farther away from 0.5. Since the middle range of probabilities is a range of values with a nearly linear CDF, the first derivative of the CDF is going to be nearly constant. As the range of values shifts away from the center, the CDF becomes less linear, indicating that that the standard deviation of the effects is larger.

Figure S.30 offers a few more simulations based on a larger range of overlapping probabilities. First, it is clear that scholars depicting substantive effects with simulation scenarios that generate probabilities near 0.5 (such as those in the [0.3, 0.7] range) will demonstrate, on average, larger effects. Second, the variability within these ranges increases as one shifts away from 0.5. Essentially, this identifies the types of political science models that are most vulnerable to high levels of variability. Models of relatively rare ([0.0, 0.4]) or likely ([0.6, 1.0]) phenomena will be much more dependent on the simulation scenario chosen because of their high standard deviations.

Tables & Figures

Figure S.1: Four Functional Forms of Duration Dependence

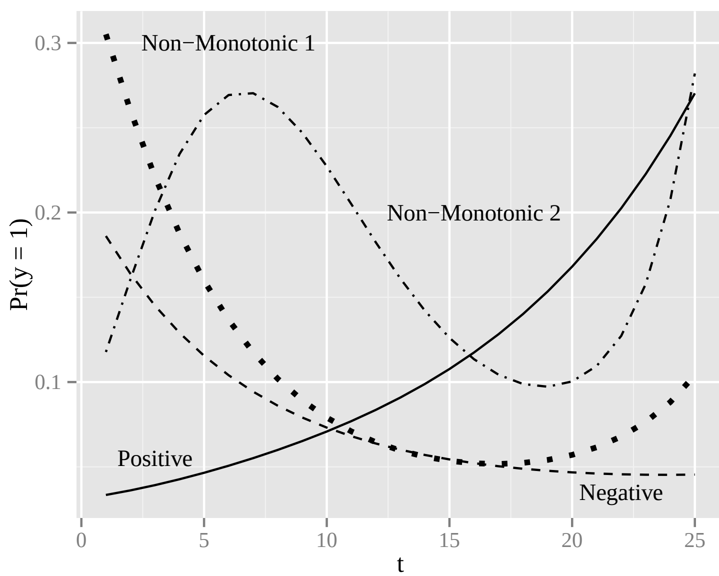
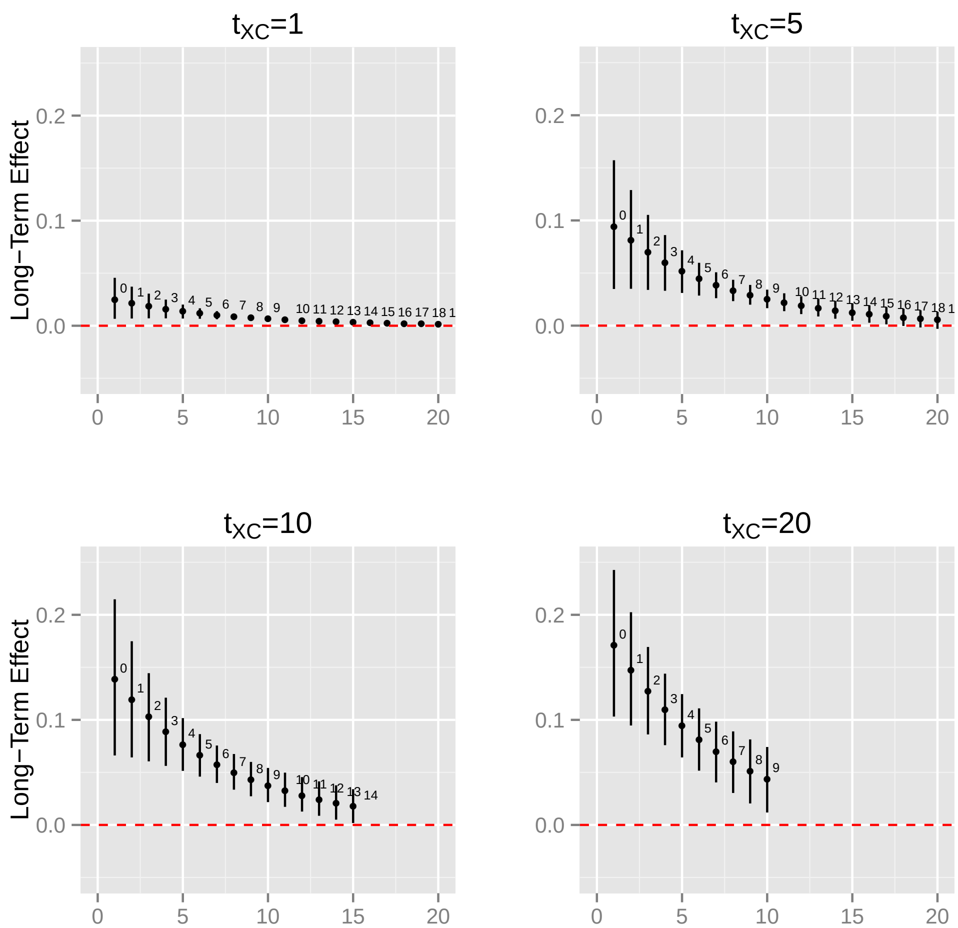
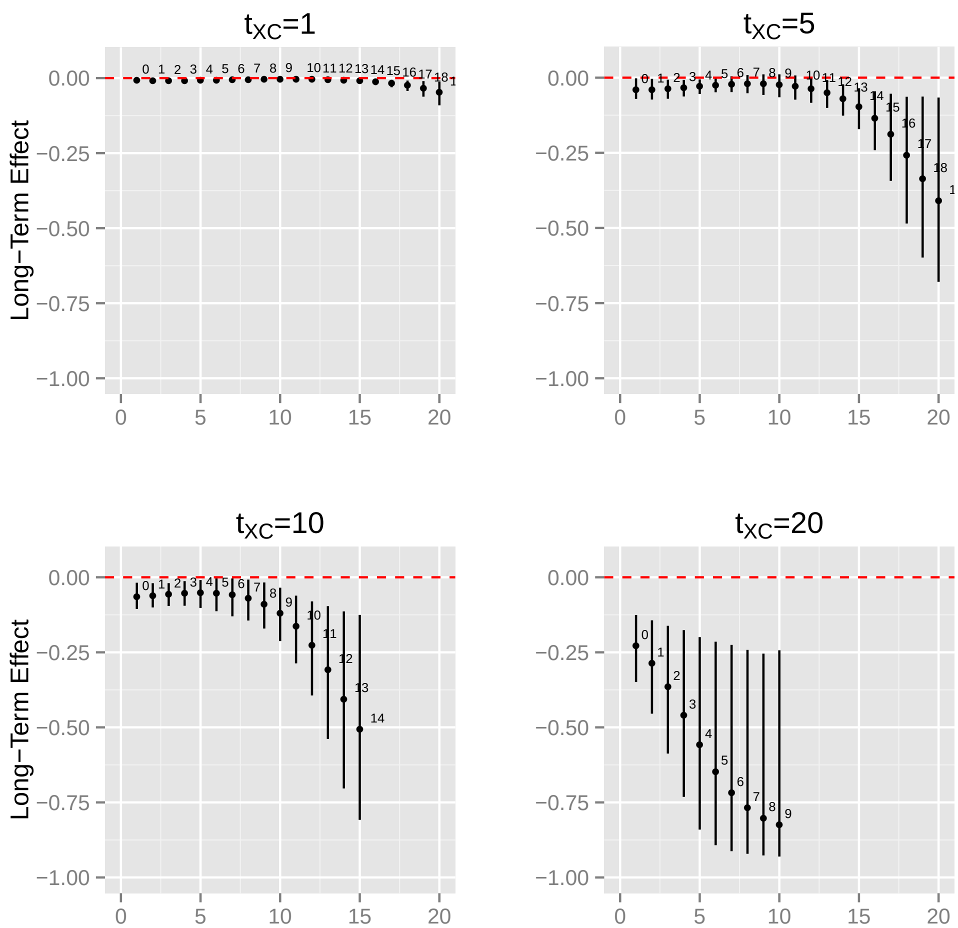


Figure S.2: Long-Term Effects for Four Scenarios for Negative Duration Dependence



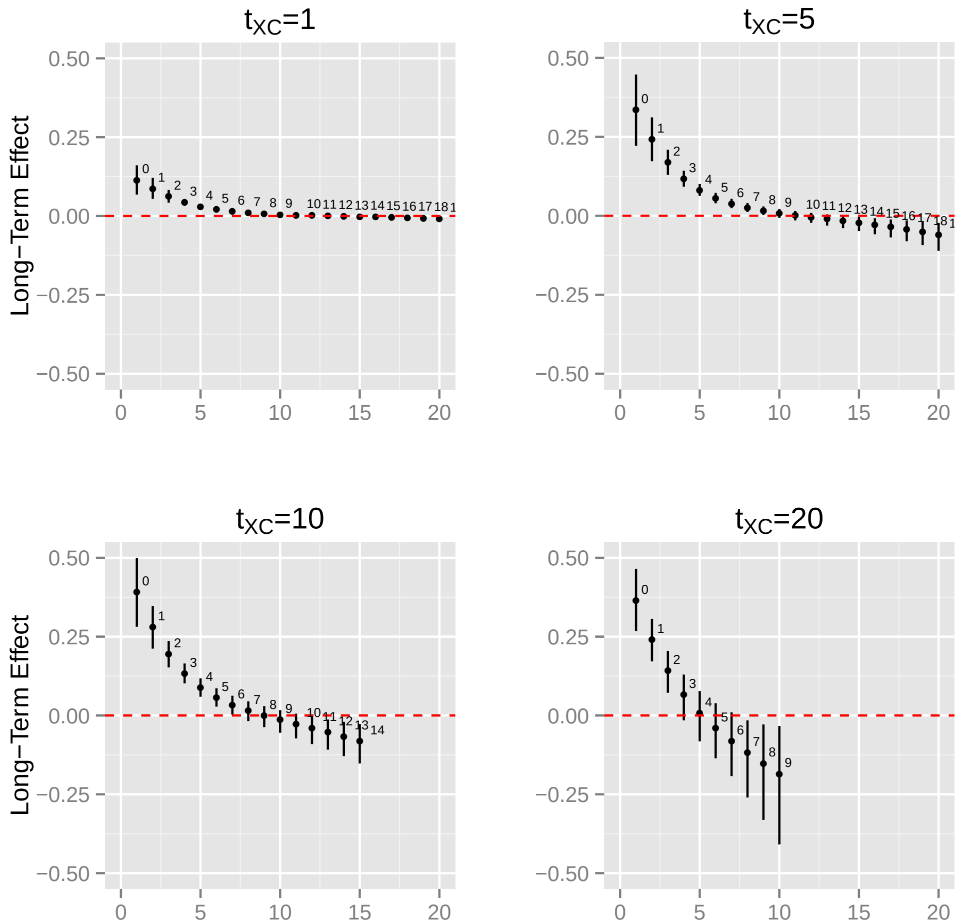
Note: The panel titles represent the values of t in the baseline scenario. The functional form is depicted in Figure S.1.

Figure S.3: Long-Term Effects for Four Scenarios for Positive Duration Dependence



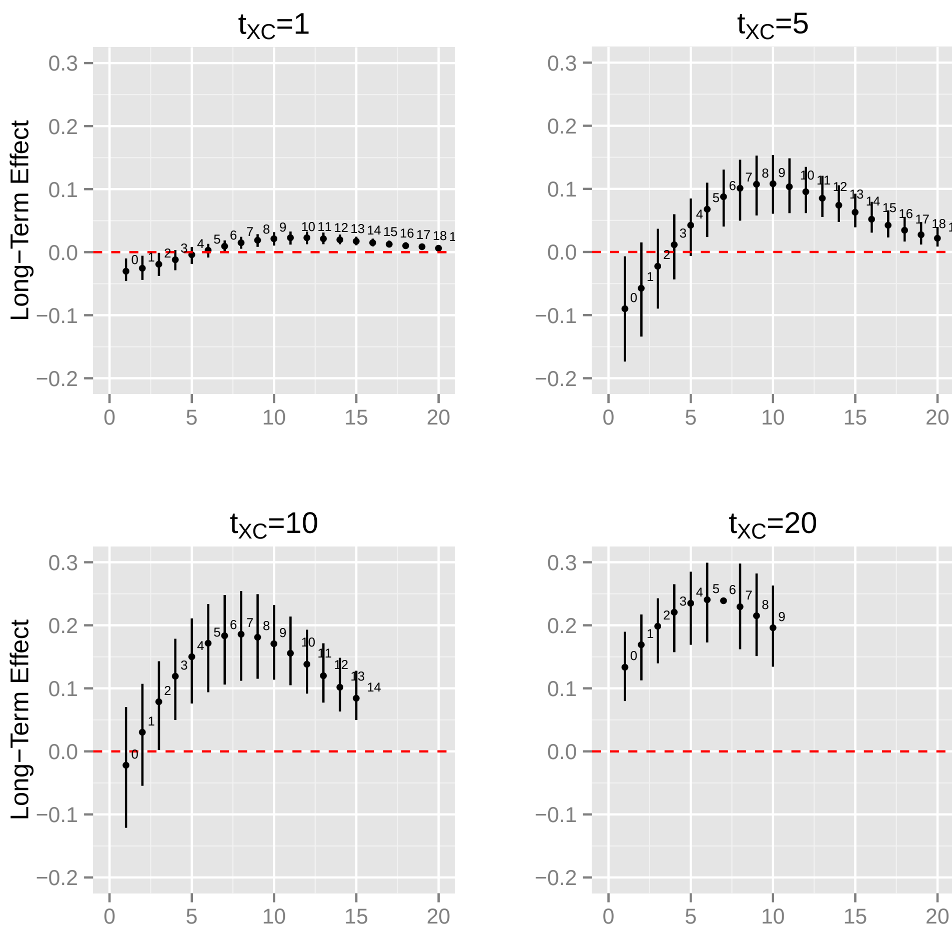
Note: The panel titles represent the values of t in the baseline scenario. The functional form is depicted in Figure S.1.

Figure S.4: Long-Term Effects for Four Scenarios for Non-Monotonic 1 Duration Dependence



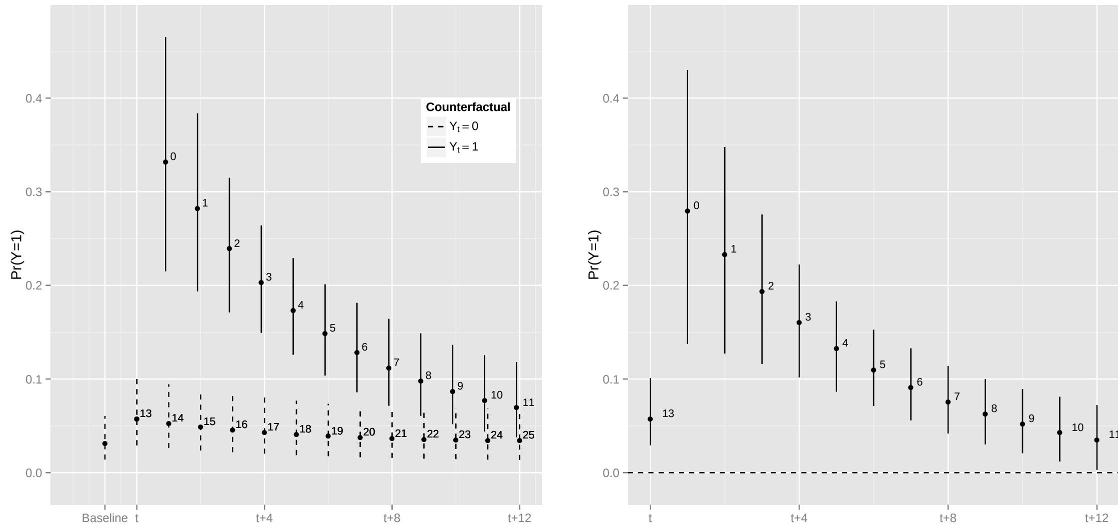
Note: The panel titles represent the values of t in the baseline scenario. The functional form is depicted in Figure S.1.

Figure S.5: Long-Term Effects for Four Scenarios for Non-Monotonic 2 Duration Dependence



Note: The panel titles represent the values of t in the baseline scenario. The functional form is depicted in Figure S.1.

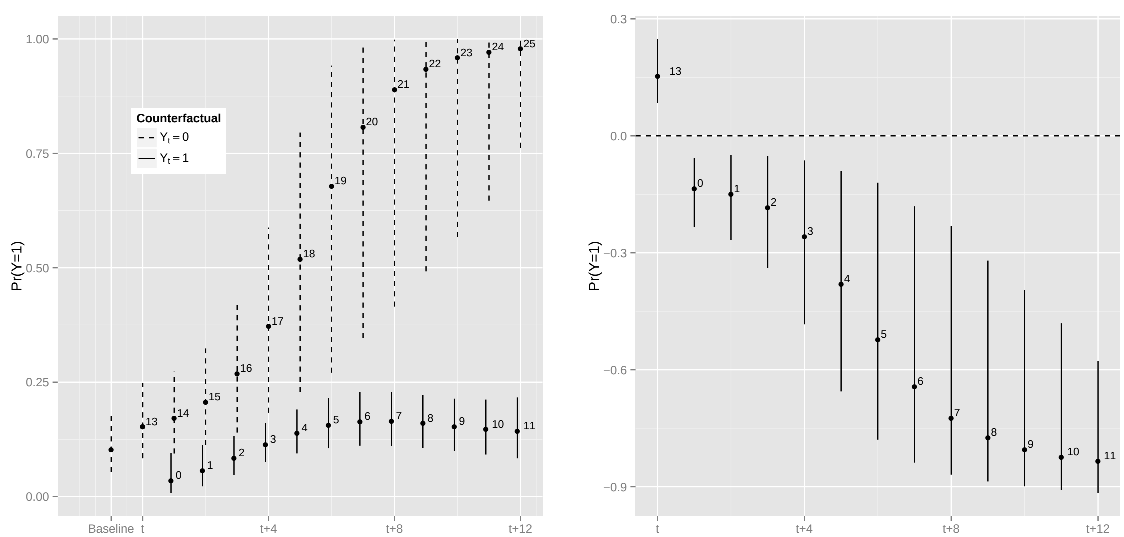
Figure S.6: Two Methods of Depicting Long-Term Effects: Negative Duration Dependence



(a) Version 1 (b) Version 2

Note: Numbers attached to points represent the values of t in the scenarios.

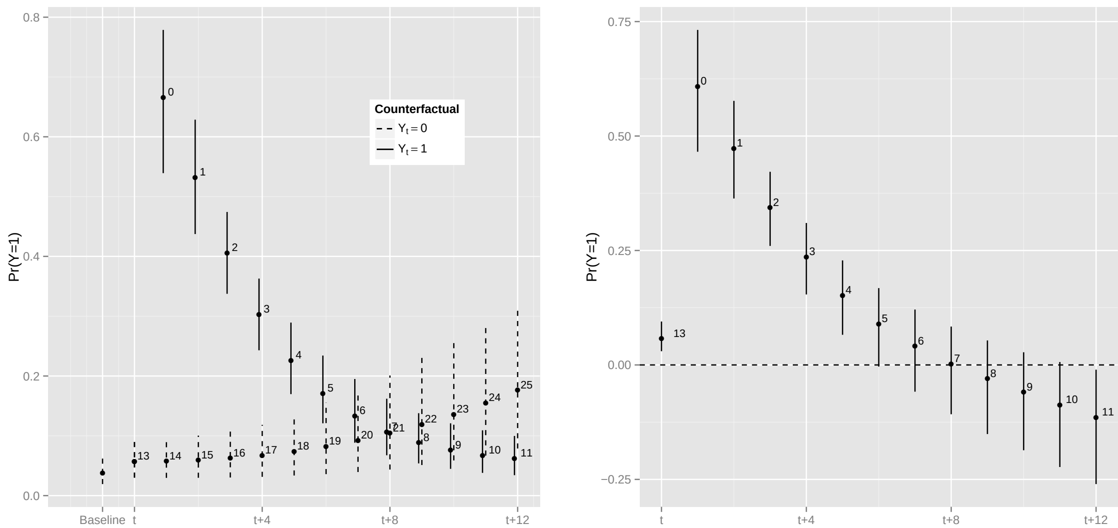
Figure S.7: Two Methods of Depicting Long-Term Effects: Positive Duration Dependence



(a) Version 1 (b) Version 2

Note: Numbers attached to points represent the values of t in the scenarios.

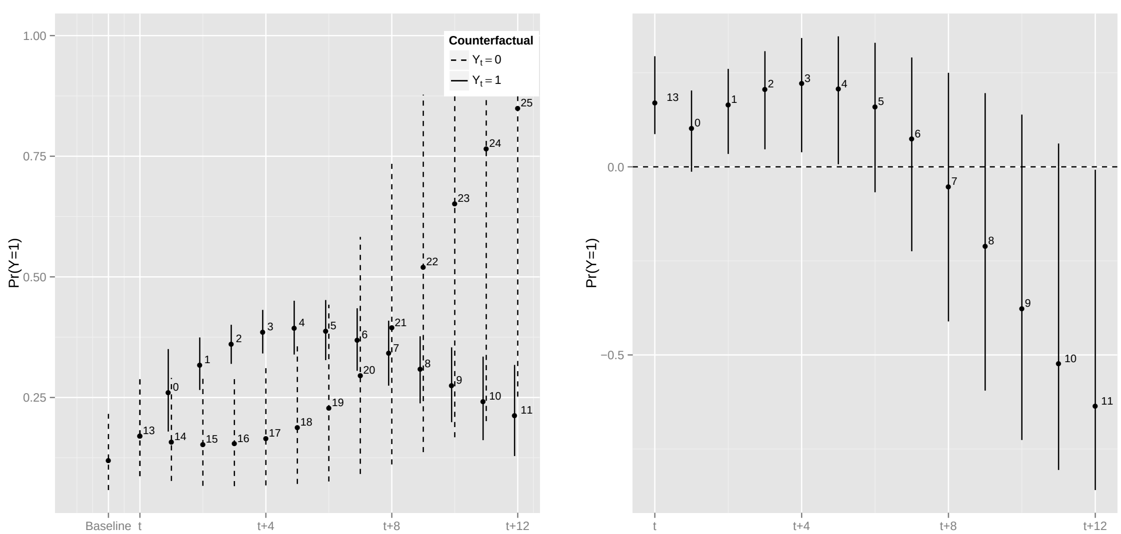
Figure S.8: Two Methods of Depicting Long-Term Effects: Non-Monotonic Duration Dependence (Parabolic)



(a) Version 1 (b) Version 2

Note: Numbers attached to points represent the values of t in the scenarios.

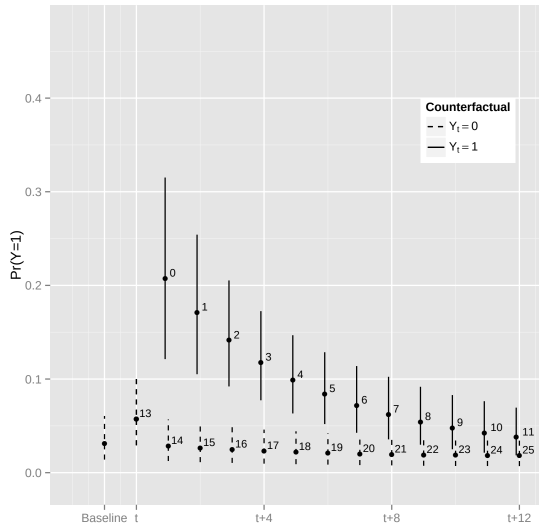
Figure S.9: Two Methods of Depicting Long-Term Effects: Non-Monotonic Duration Dependence (Log-Logistic)



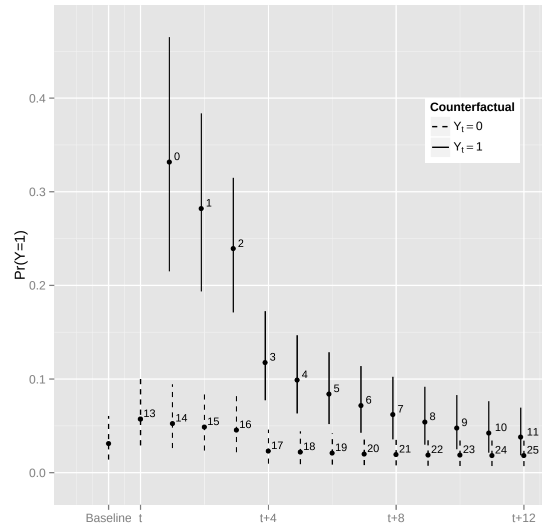
(a) Version 1 (b) Version 2

Note: Numbers attached to points represent the values of t in the scenarios.

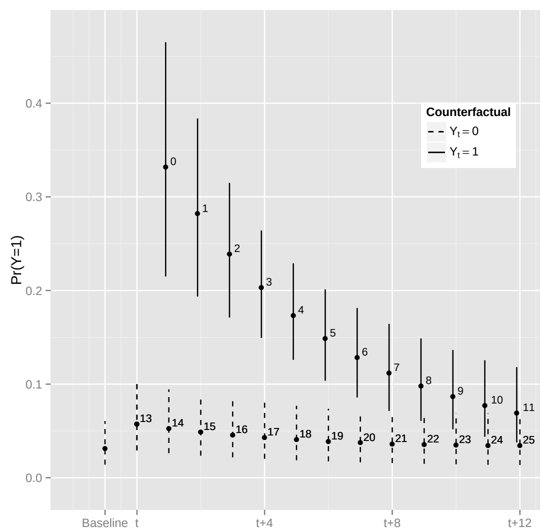
Figure S.10: Variations of Long-Term Effects Based on the Counterfactual of Interest for Negative Duration Dependence



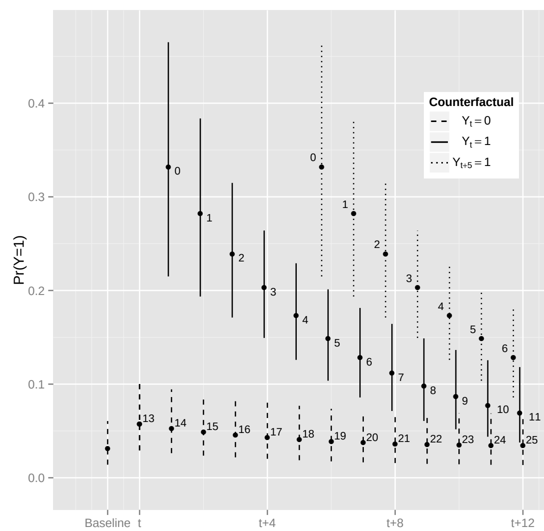
(a) One-Time



(b) Temporary



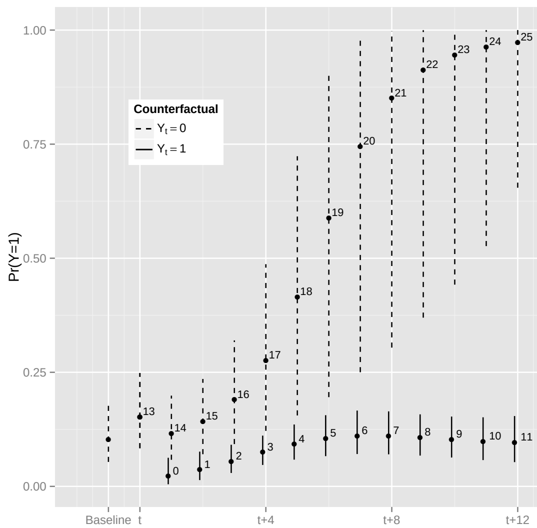
(c) Permanent



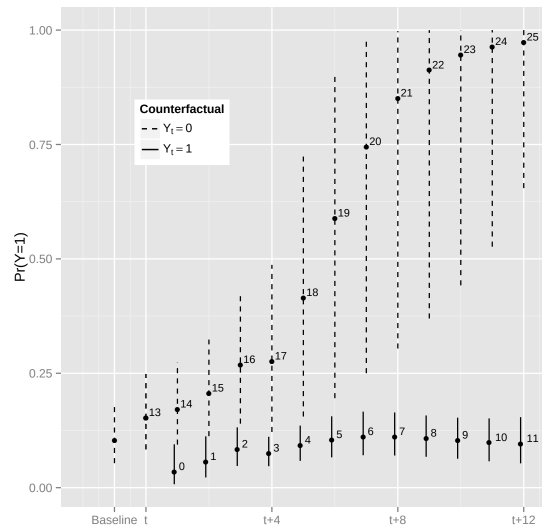
(d) Compounded

Note: Numbers attached to points represent the values of t in the scenarios.

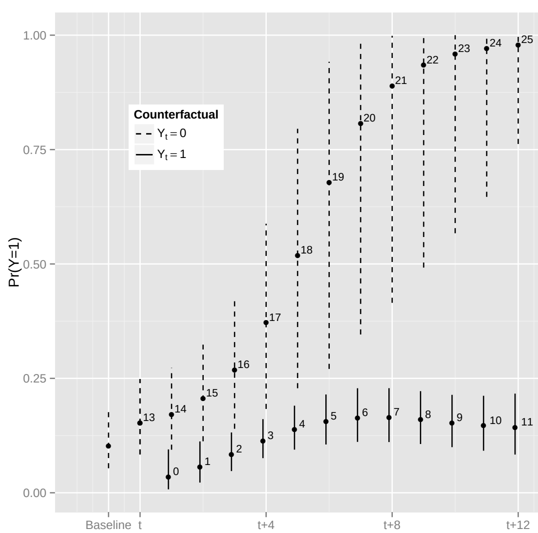
Figure S.11: Variations of Long-Term Effects Based on the Counterfactual of Interest for Positive Duration Dependence



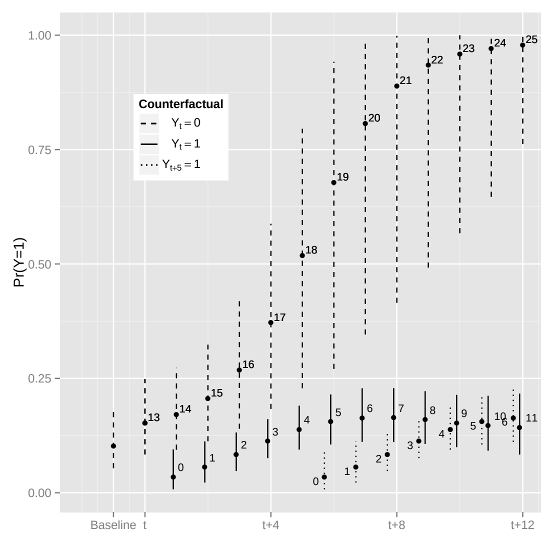
(a) One-Time



(b) Temporary



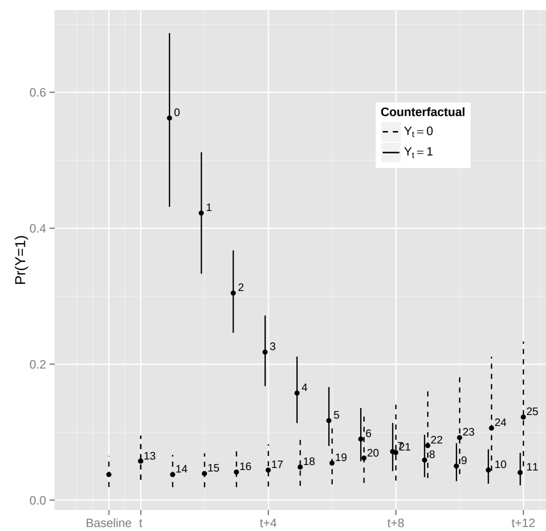
(c) Permanent



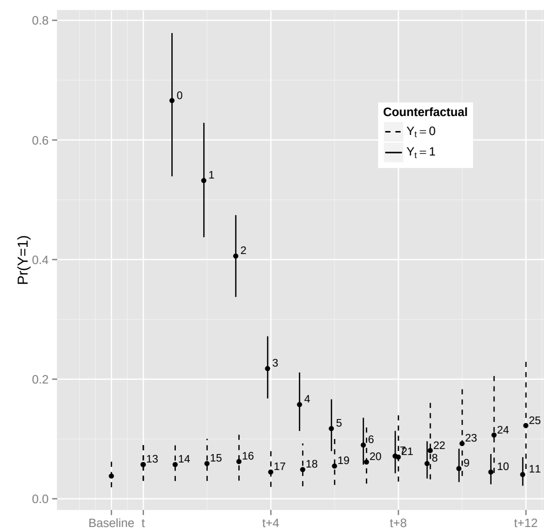
(d) Compounded

Note: Numbers attached to points represent the values of t in the scenarios.

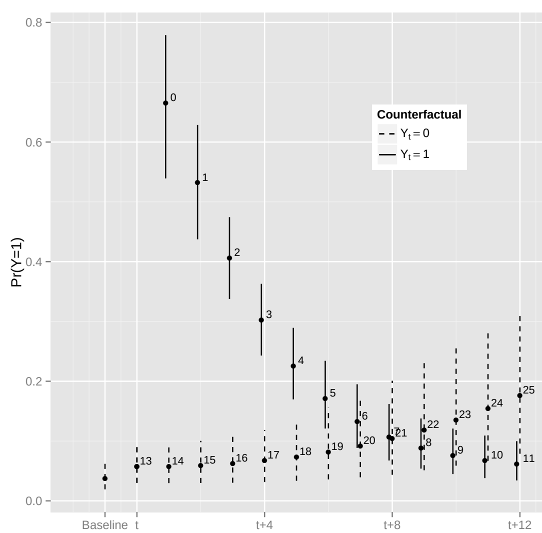
Figure S.12: Variations of Long-Term Effects Based on the Counterfactual of Interest for Non-Monotonic Duration Dependence (Parabolic)



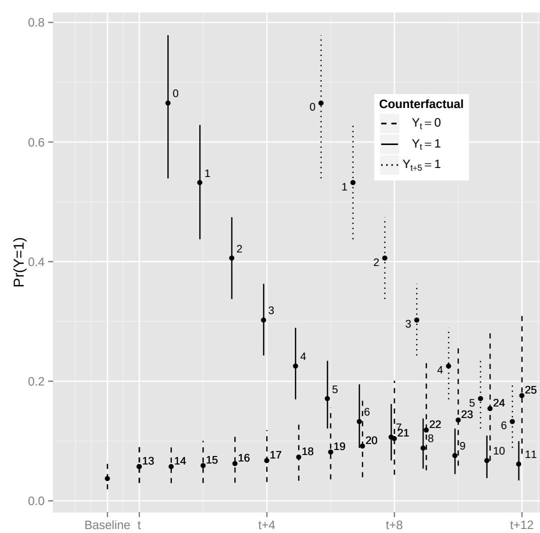
(a) One-Time



(b) Temporary



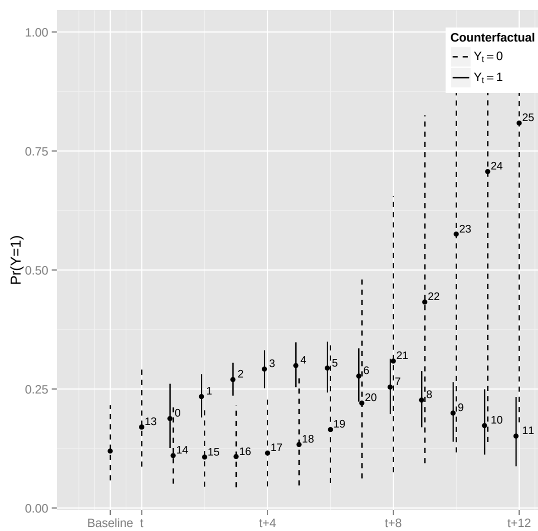
(c) Permanent



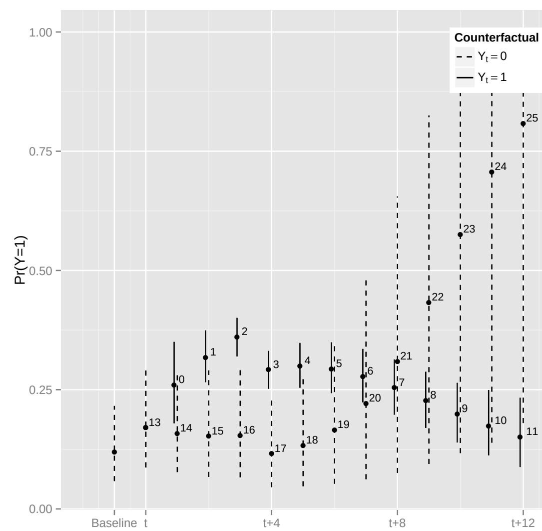
(d) Compounded

Note: Numbers attached to points represent the values of t in the scenarios.

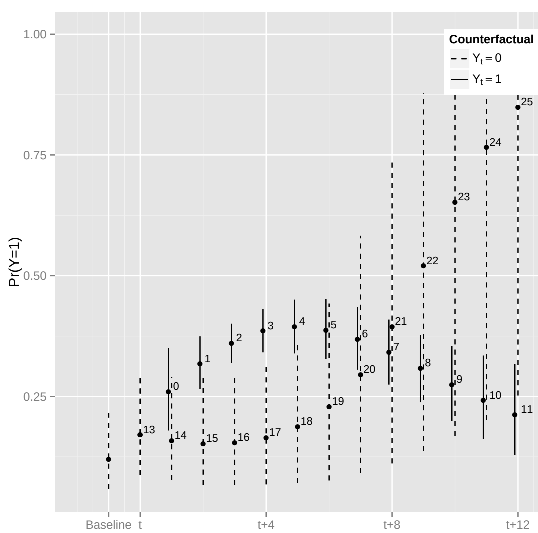
Figure S.13: Variations of Long-Term Effects Based on the Counterfactual of Interest for Non-Monotonic Duration Dependence (Log-Logistic)



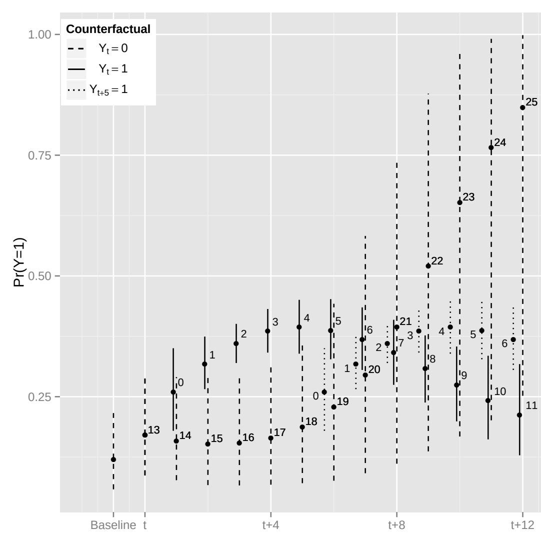
(a) One-Time



(b) Temporary



(c) Permanent



(d) Compounded

Note: Numbers attached to points represent the values of t in the scenarios.

Table S.1: Performance of β_x under Various Circumstances in Monte Carlo Experiments: Negative Duration Dependence

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

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

Table S.2: Performance of β_x under Various Circumstances in Monte Carlo Experiments: Positive Duration Dependence

Scenario	**Avg.** β_{X_K}	**Bias**	**MSE**	**SE**	**SD**
Exponential (Flat)					
N=1000; 1s=12%	0.974	0.090	0.013	0.114	0.110
N=1000; 1s=20%	0.990	0.074	0.008	0.090	0.091
N=5000; 1s=12%	0.968	0.048	0.004	0.051	0.051
N=5000; 1s=20%	0.987	0.034	0.002	0.040	0.040
N=10,000; 1s=12%	0.967	0.041	0.002	0.036	0.037
N=10,000; 1s=20%	0.986	0.026	0.001	0.028	0.029
Temporal Dummies					
N=1000; 1s=12%	1.010	0.089	0.013	0.118	0.113
N=1000; 1s=20%	1.006	0.074	0.009	0.092	0.093
N=5000; 1s=12%	1.002	0.041	0.003	0.052	0.052
N=5000; 1s=20%	1.001	0.032	0.002	0.041	0.040
N=10,000; 1s=12%	1.000	0.030	0.001	0.037	0.038
N=10,000; 1s=20%	1.000	0.023	0.001	0.029	0.029
Cubic Polynomials					
N=1000; 1s=12%	1.014	0.090	0.013	0.118	0.113
N=1000; 1s=20%	1.009	0.074	0.009	0.092	0.093
N=5000; 1s=12%	1.003	0.041	0.003	0.052	0.052
N=5000; 1s=20%	1.001	0.032	0.002	0.041	0.040
N=10,000; 1s=12%	1.002	0.030	0.001	0.037	0.038
N=10,000; 1s=20%	1.000	0.023	0.001	0.029	0.029
B-Splines					
N=1000; 1s=12%	1.013	0.090	0.013	0.118	0.113
N=1000; 1s=20%	1.009	0.074	0.009	0.092	0.093
N=5000; 1s=12%	1.003	0.041	0.003	0.052	0.052
N=5000; 1s=20%	1.001	0.032	0.002	0.041	0.040
N=10,000; 1s=12%	1.002	0.030	0.001	0.037	0.038
N=10,000; 1s=20%	1.000	0.023	0.001	0.029	0.029
Automated Smoothing Splines					
N=1000; 1s=12%	1.009	0.089	0.013	0.118	0.113
N=1000; 1s=20%	1.006	0.074	0.009	0.092	0.093
N=5000; 1s=12%	1.002	0.041	0.003	0.052	0.052
N=5000; 1s=20%	1.001	0.032	0.002	0.041	0.040
N=10,000; 1s=12%	1.000	0.030	0.001	0.037	0.038
N=10,000; 1s=20%	1.000	0.023	0.001	0.029	0.029

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

Table S.3: Performance of β_x under Various Circumstances in Monte Carlo Experiments: Non-Monotonic (Parabolic) Duration Dependence

Scenario	**Avg.** β_{X_K}	**Bias**	**MSE**	**SE**	**SD**
Exponential (Flat)					
N=1000; 1s=24%	0.913	0.100	0.014	0.081	0.080
N=1000; 1s=50%	0.951	0.070	0.007	0.069	0.070
N=5000; 1s=24%	0.914	0.086	0.009	0.036	0.038
N=5000; 1s=50%	0.952	0.050	0.003	0.031	0.032
N=10,000; 1s=24%	0.912	0.088	0.008	0.026	0.026
N=10,000; 1s=50%	0.949	0.051	0.003	0.022	0.021
Temporal Dummies					
N=1000; 1s=24%	0.998	0.070	0.008	0.087	0.087
N=1000; 1s=50%	1.001	0.057	0.005	0.072	0.070
N=5000; 1s=24%	1.000	0.032	0.002	0.039	0.040
N=5000; 1s=50%	1.002	0.026	0.001	0.032	0.032
N=10,000; 1s=24%	1.000	0.022	0.001	0.027	0.028
N=10,000; 1s=50%	0.999	0.018	0.001	0.023	0.022
Cubic Polynomials					
N=1000; 1s=24%	1.003	0.070	0.008	0.087	0.088
N=1000; 1s=50%	1.003	0.057	0.005	0.073	0.070
N=5000; 1s=24%	1.003	0.032	0.002	0.039	0.040
N=5000; 1s=50%	1.003	0.026	0.001	0.032	0.032
N=10,000; 1s=24%	1.001	0.022	0.001	0.027	0.028
N=10,000; 1s=50%	1.000	0.018	0.001	0.023	0.022
B-Splines					
N=1000; 1s=24%	1.003	0.070	0.008	0.087	0.088
N=1000; 1s=50%	1.003	0.057	0.005	0.073	0.071
N=5000; 1s=24%	1.002	0.032	0.002	0.039	0.040
N=5000; 1s=50%	1.003	0.026	0.001	0.032	0.032
N=10,000; 1s=24%	1.000	0.022	0.001	0.027	0.028
N=10,000; 1s=50%	1.000	0.018	0.001	0.023	0.022
Automated Smoothing Splines					
N=1000; 1s=24%	1.000	0.070	0.008	0.087	0.088
N=1000; 1s=50%	1.000	0.057	0.005	0.073	0.070
N=5000; 1s=24%	1.002	0.032	0.002	0.039	0.040
N=5000; 1s=50%	1.002	0.026	0.001	0.033	0.032
N=10,000; 1s=24%	1.000	0.022	0.001	0.027	0.028
N=10,000; 1s=50%	1.000	0.018	0.001	0.023	0.022

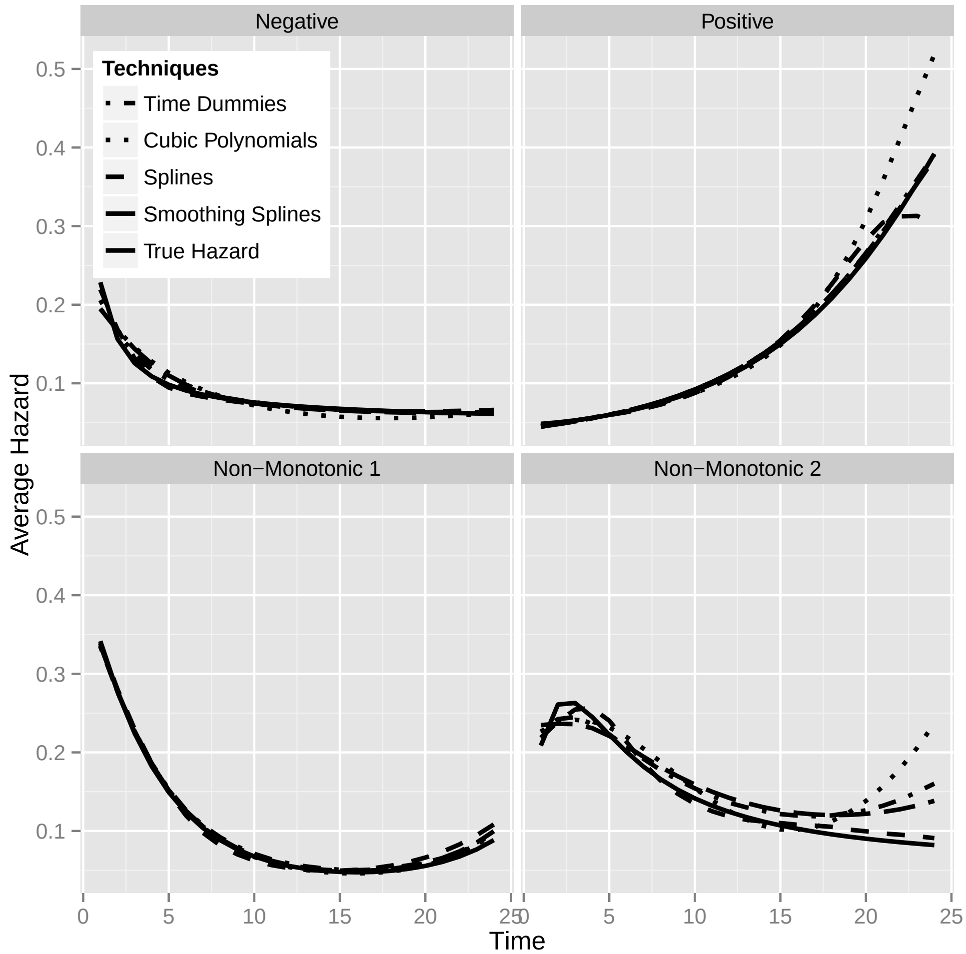
Note: **Bias** is the mean of absolute bias: $|\hat{\beta}_x - 1|$. **MSE** is the mean of expected squared bias: $E[(\hat{\beta}_x - 1)^2]$. **SE** is the mean of the simulated standard errors. **SD** is the standard deviation of the estimates.

Table S.4: Performance of β_x under Various Circumstances in Monte Carlo Experiments: Non-Monotonic (Log-Logistic) Duration Dependence

Scenario	**Avg.** β_{X_K}	**Bias**	**MSE**	**SE**	**SD**
Exponential (Flat)					
N=1000; 1s=27%	0.986	0.065	0.007	0.081	0.080
N=1000; 1s=46%	0.998	0.057	0.005	0.071	0.070
N=5000; 1s=27%	0.986	0.032	0.002	0.036	0.037
N=5000; 1s=46%	0.996	0.026	0.001	0.032	0.032
N=10,000; 1s=27%	0.984	0.025	0.001	0.026	0.026
N=10,000; 1s=46%	0.994	0.018	0.001	0.022	0.022
Temporal Dummies					
N=1000; 1s=27%	1.002	0.065	0.007	0.082	0.081
N=1000; 1s=46%	1.005	0.057	0.005	0.072	0.070
N=5000; 1s=27%	1.002	0.030	0.001	0.037	0.037
N=5000; 1s=46%	1.002	0.026	0.001	0.032	0.032
N=10,000; 1s=27%	1.000	0.022	0.001	0.026	0.027
N=10,000; 1s=46%	0.999	0.018	0.001	0.023	0.022
Cubic Polynomials					
N=1000; 1s=27%	1.001	0.064	0.007	0.082	0.081
N=1000; 1s=46%	1.007	0.057	0.005	0.072	0.070
N=5000; 1s=27%	0.998	0.030	0.001	0.037	0.037
N=5000; 1s=46%	1.002	0.025	0.001	0.032	0.032
N=10,000; 1s=27%	0.996	0.022	0.001	0.026	0.027
N=10,000; 1s=46%	0.999	0.018	0.001	0.023	0.022
B-Splines					
N=1000; 1s=27%	1.003	0.065	0.007	0.082	0.081
N=1000; 1s=46%	1.006	0.057	0.005	0.072	0.070
N=5000; 1s=27%	1.000	0.030	0.001	0.037	0.037
N=5000; 1s=46%	1.002	0.025	0.001	0.032	0.032
N=10,000; 1s=27%	0.999	0.022	0.001	0.026	0.027
N=10,000; 1s=46%	0.999	0.018	0.001	0.023	0.022
Automated Smoothing Splines					
N=1000; 1s=27%	0.999	0.064	0.007	0.082	0.081
N=1000; 1s=46%	1.003	0.057	0.005	0.072	0.070
N=5000; 1s=27%	0.999	0.030	0.001	0.037	0.037
N=5000; 1s=46%	1.001	0.025	0.001	0.032	0.032
N=10,000; 1s=27%	0.997	0.022	0.001	0.026	0.027
N=10,000; 1s=46%	0.999	0.018	0.001	0.023	0.022

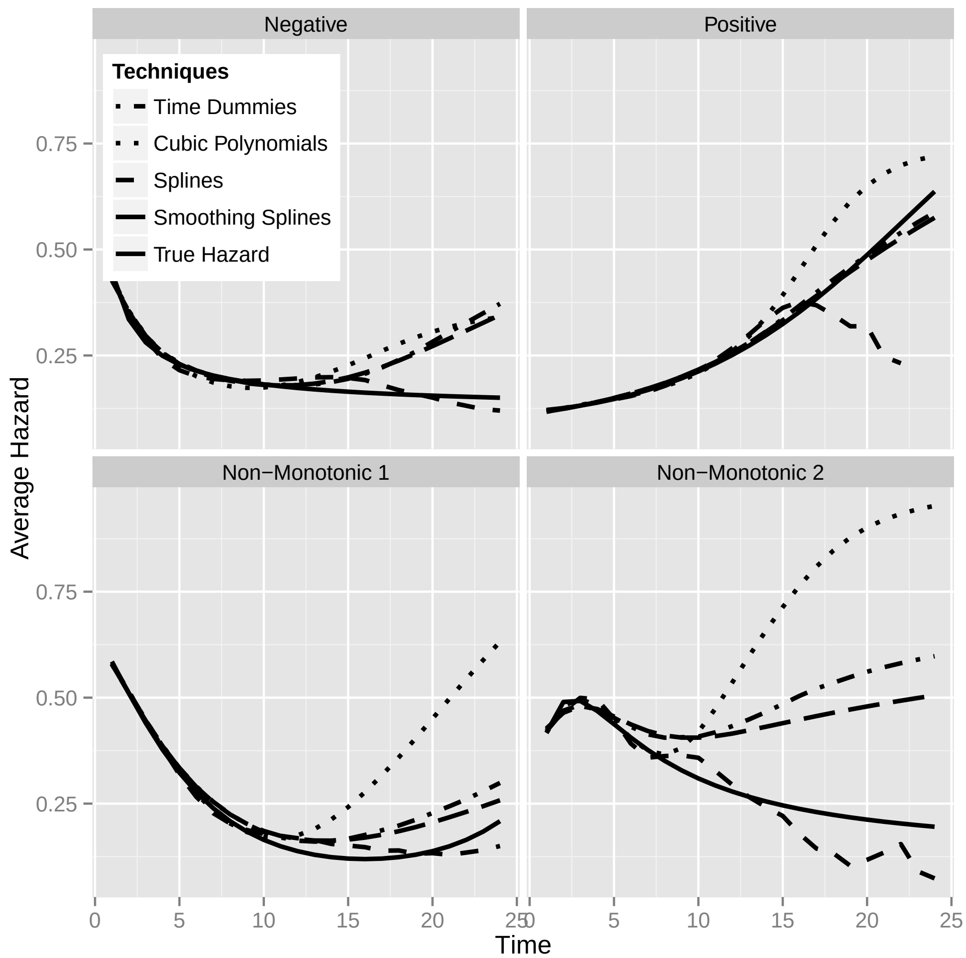
Note: **Bias** is the mean of absolute bias: $|\hat{\beta}_x - 1|$. **MSE** is the mean of expected squared bias: $E[(\hat{\beta}_x - 1)^2]$. **SE** is the mean of the simulated standard errors. **SD** is the standard deviation of the estimates.

Figure S.14: Average Hazard Rates Compared to True Hazard Rates



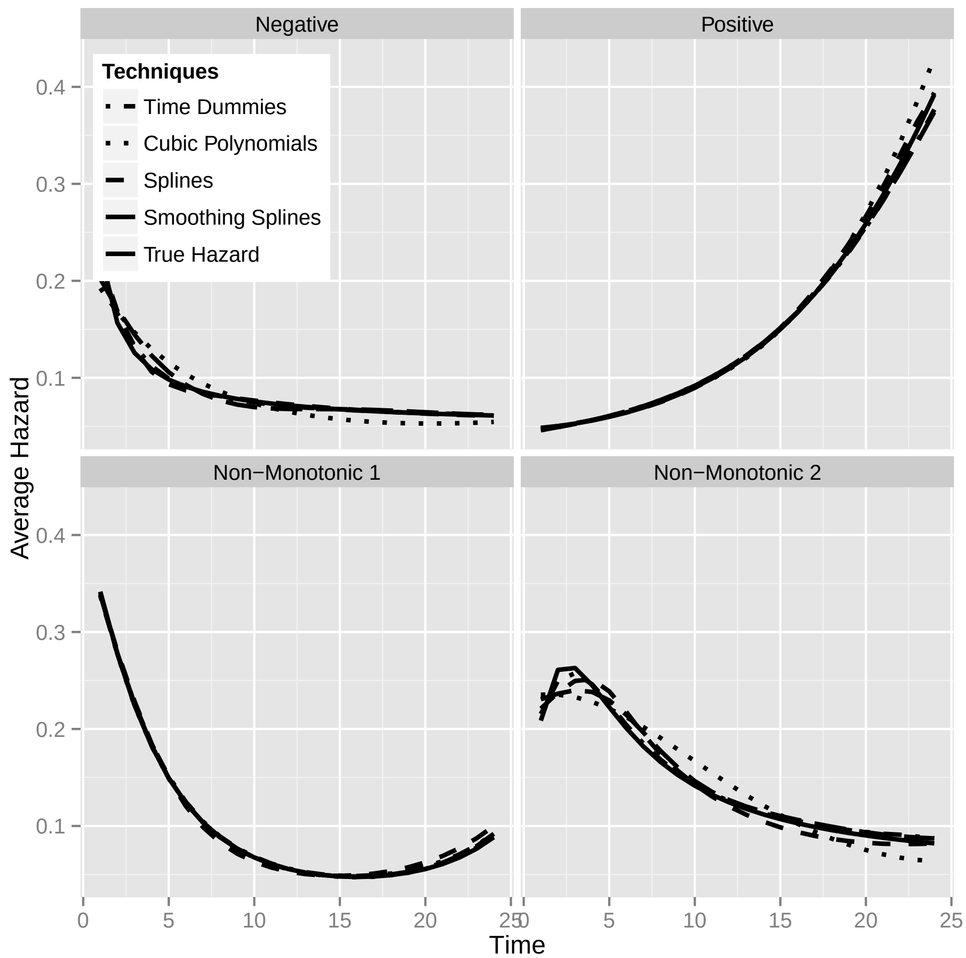
Note: N=1,000, sims = 1,000, beta = c(-3, 1).

Figure S.15: Average Hazard Rates Compared to True Hazard Rates



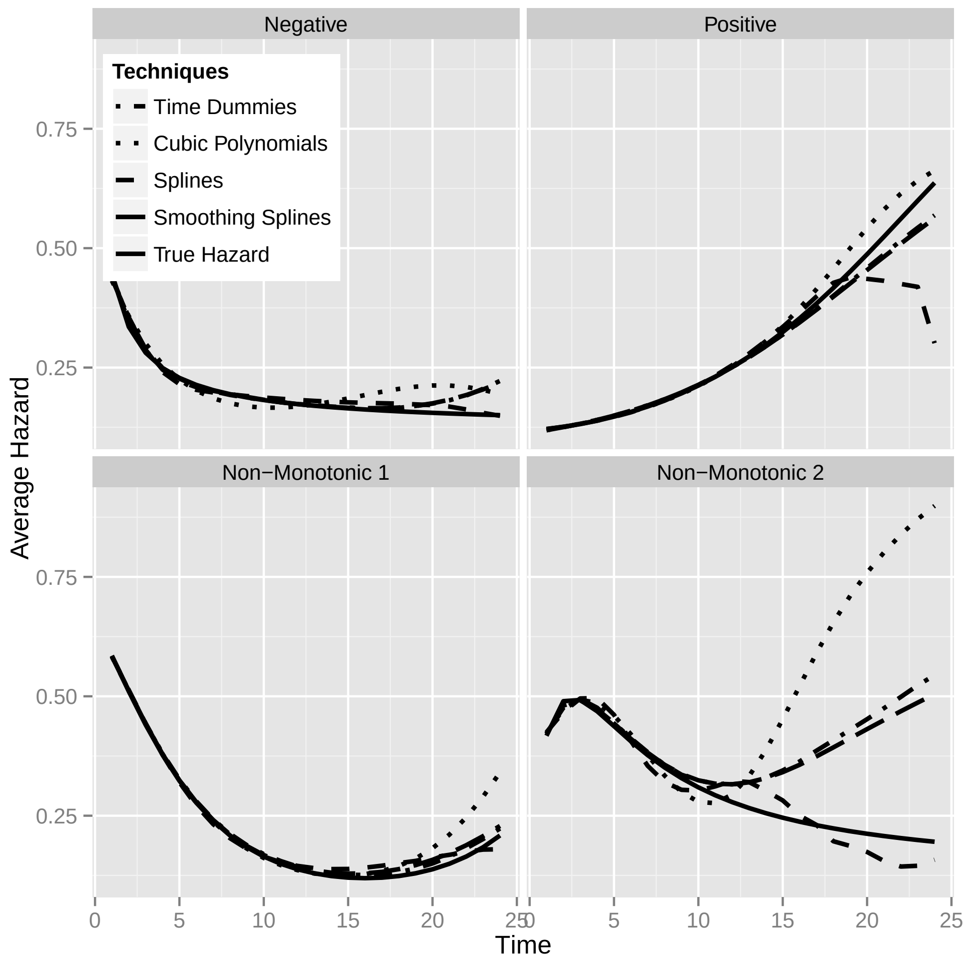
Note: N=1,000, sims = 1,000, beta = c(-2, 1).

Figure S.16: Average Hazard Rates Compared to True Hazard Rates



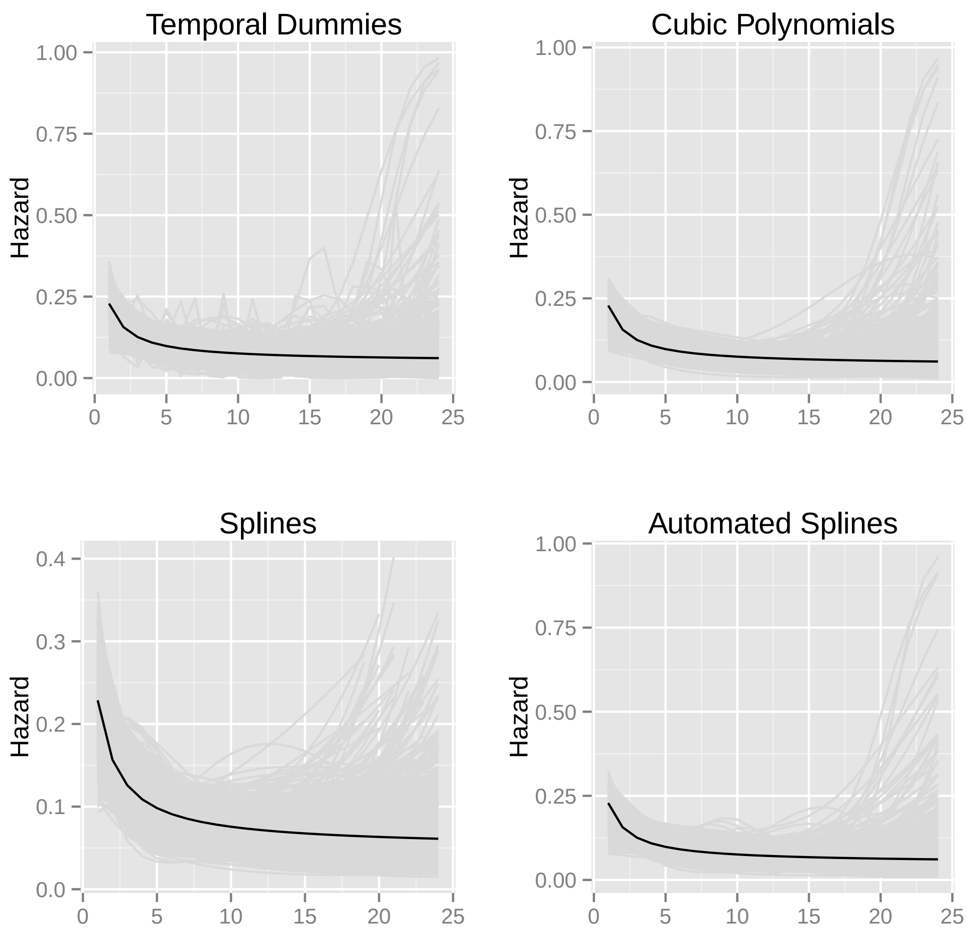
Note: N=5,000, sims = 1,000, beta = c(-3, 1).

Figure S.17: Average Hazard Rates Compared to True Hazard Rates



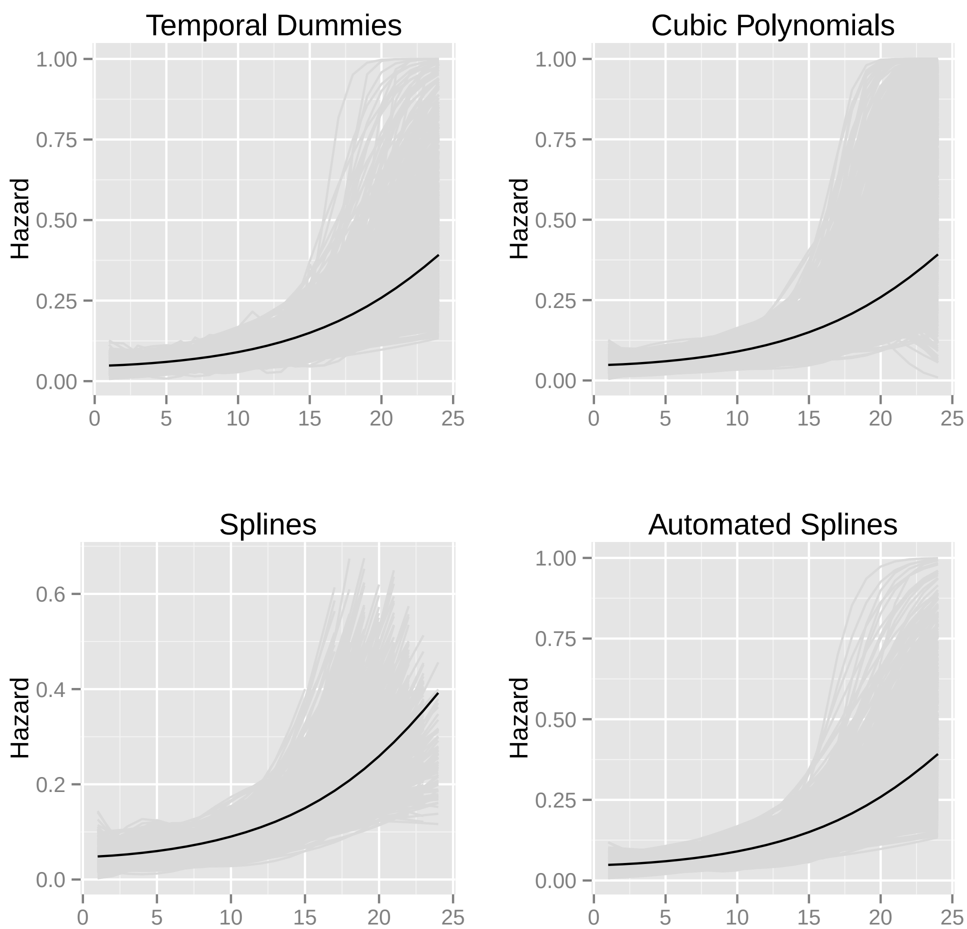
Note: N=5,000, sims = 1,000, beta = c(-2, 1).

Figure S.18: Estimated Hazard Rates Compared to True Hazard Rates: Negative Duration Dependence



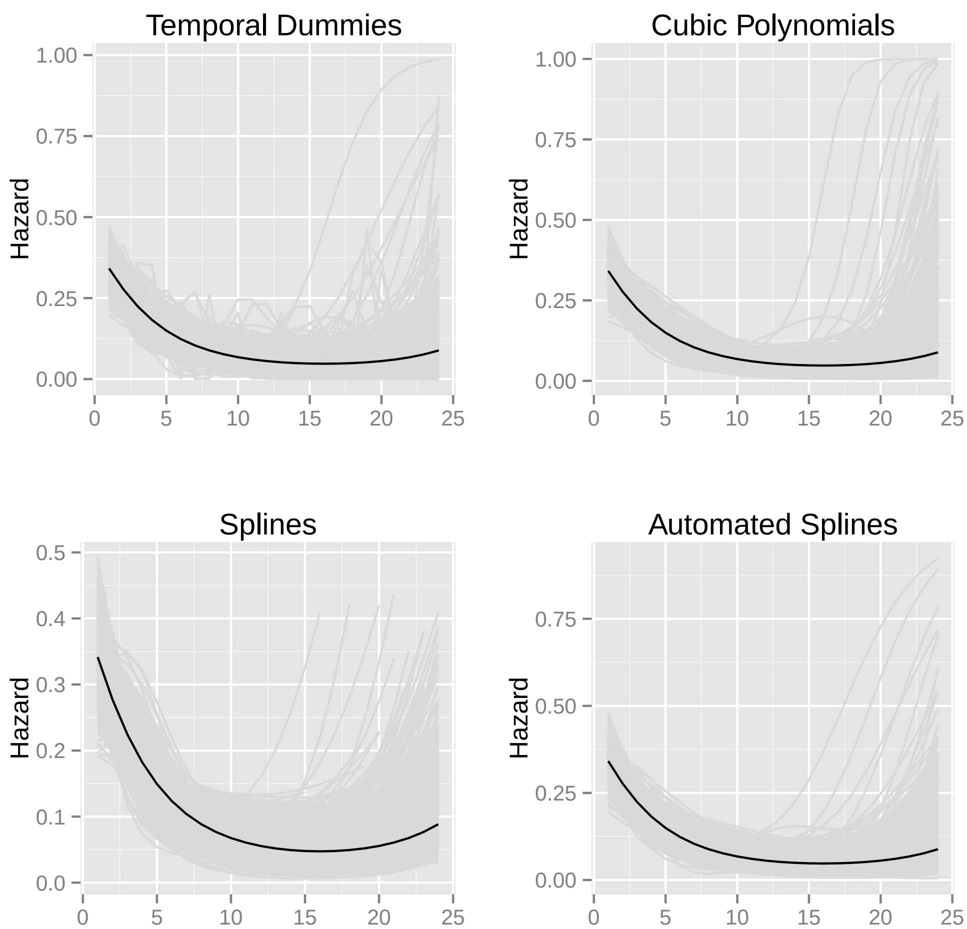
Note: N=1,000, sims = 1,000, beta = c(-3, 1).

Figure S.19: Estimated Hazard Rates Compared to True Hazard Rates: Positive Duration Dependence



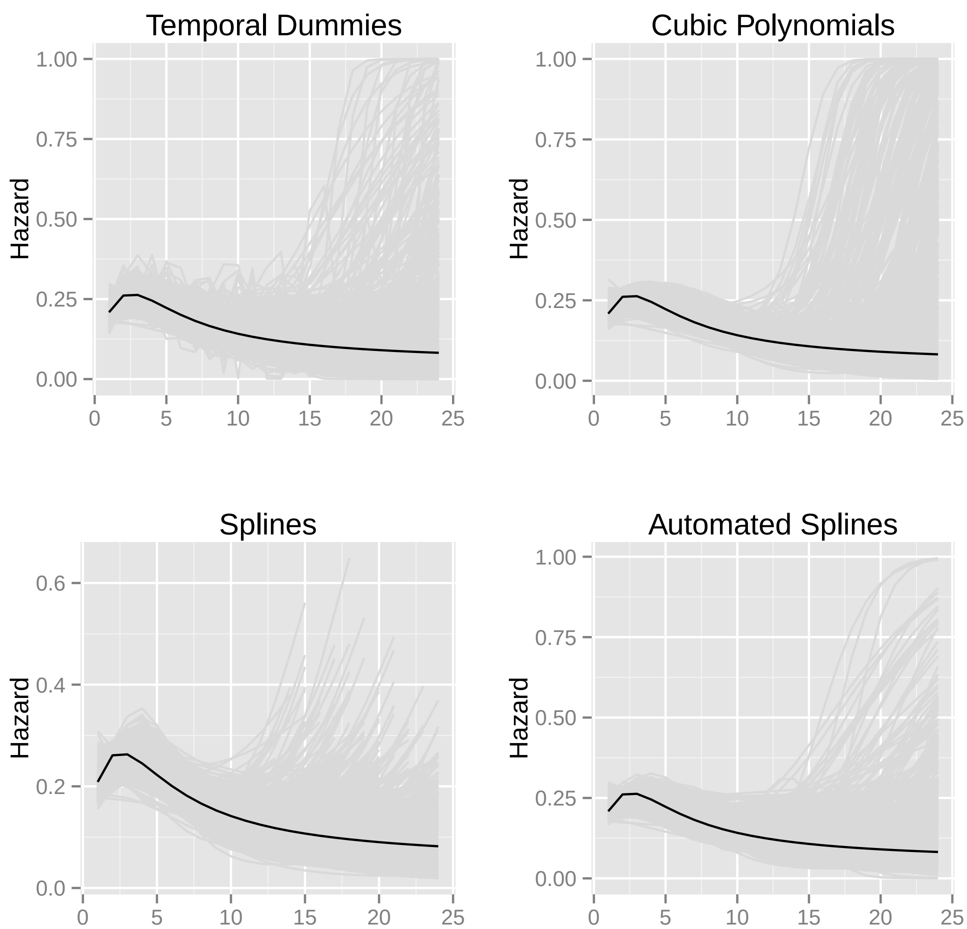
Note: N=1,000, sims = 1,000, beta = c(-3, 1).

Figure S.20: Estimated Hazard Rates Compared to True Hazard Rates: Non-Monotonic Duration Dependence (Parabolic)



Note: N=1,000, sims = 1,000, beta = c(-3, 1).

Figure S.21: Estimated Hazard Rates Compared to True Hazard Rates: Non-Monotonic Duration Dependence (Log-Logistic)



Note: N=1,000, sims = 1,000, beta = c(-3, 1).

Table S.5: Performance of β_x under Various Circumstances in Monte Carlo Experiments: Negative Duration Dependence with Positive Non-Proportional Hazards

Scenario	**Avg.** β_{X_K}	**Bias**	**MSE**	**SE**	**SD**
Temporal Dummies					
$\beta_{Xt} = 0.2$	1.001	0.063	0.006	0.079	0.079
$\beta_{Xt} = 0.1$	1.006	0.066	0.007	0.085	0.083
$\beta_{Xt} = 0.04$	1.006	0.077	0.010	0.092	0.099
$\beta_{Xt} = 0.02$	1.003	0.076	0.009	0.096	0.096
Cubic Polynomials					
$\beta_{Xt} = 0.2$	1.004	0.063	0.006	0.080	0.079
$\beta_{Xt} = 0.1$	1.008	0.066	0.007	0.085	0.083
$\beta_{Xt} = 0.04$	1.006	0.077	0.010	0.092	0.098
$\beta_{Xt} = 0.02$	1.003	0.075	0.009	0.096	0.095
B-Splines					
$\beta_{Xt} = 0.2$	1.004	0.063	0.006	0.080	0.079
$\beta_{Xt} = 0.1$	1.008	0.066	0.007	0.085	0.083
$\beta_{Xt} = 0.04$	1.008	0.077	0.010	0.092	0.098
$\beta_{Xt} = 0.02$	1.006	0.075	0.009	0.096	0.096
Automated Smoothing Splines					
$\beta_{Xt} = 0.2$	1.000	0.063	0.006	0.079	0.079
$\beta_{Xt} = 0.1$	1.005	0.066	0.007	0.085	0.083
$\beta_{Xt} = 0.04$	1.004	0.077	0.010	0.092	0.097
$\beta_{Xt} = 0.02$	1.001	0.075	0.009	0.095	0.095

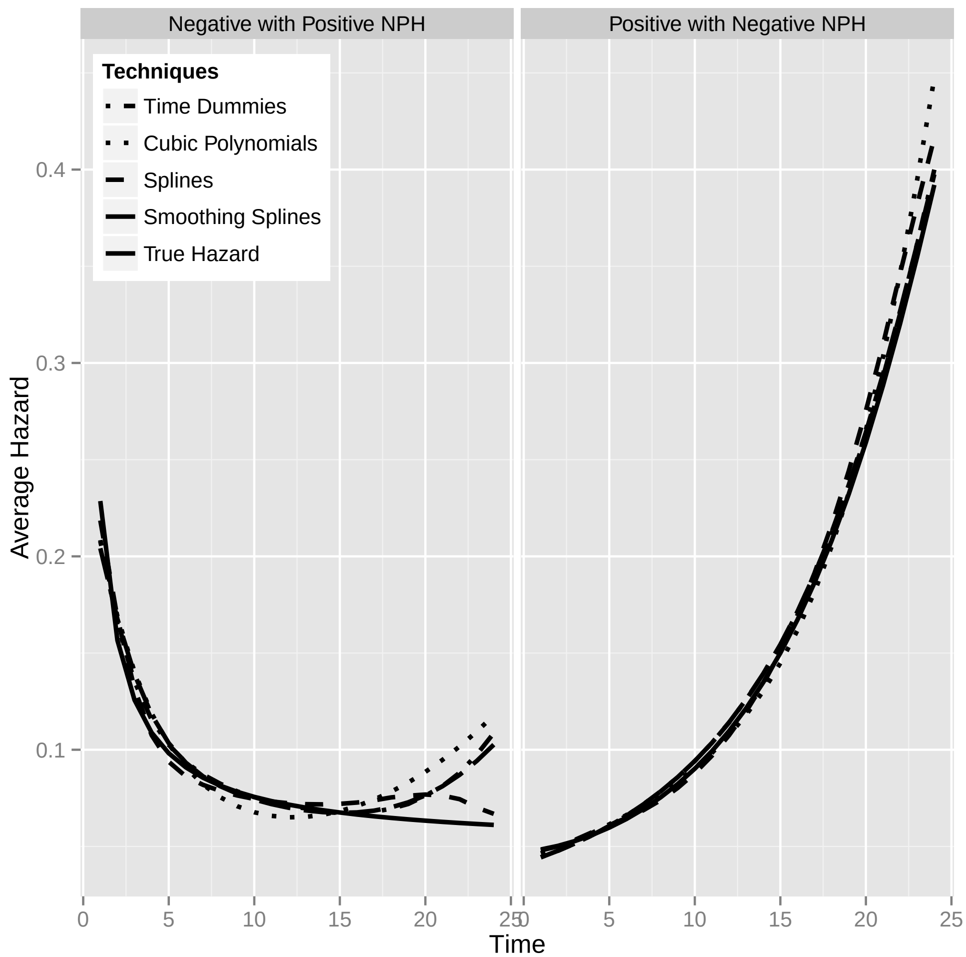
Note: N=1000, betas = c(-3,1), x is drawn randomly from a uniform distribution, $\in [-2, 2]$. **Bias** is the mean of absolute bias: $|\hat{\beta}_x - 1|$. **MSE** is the mean of expected squared bias: $E[(\hat{\beta}_x - 1)^2]$. **SE** is the mean of the simulated standard errors. **SD** is the standard deviation of the estimates.

Table S.6: Performance of β_x under Various Circumstances in Monte Carlo Experiments: Positive Duration Dependence with Negative Non-Proportional Hazards

Scenario	**Avg.** β_{X_K}	**Bias**	**MSE**	**SE**	**SD**
Temporal Dummies					
$\beta_{Xt} = -0.2$	1.036	0.206	0.073	0.263	0.268
$\beta_{Xt} = -0.1$	1.014	0.121	0.023	0.148	0.152
$\beta_{Xt} = -0.04$	1.006	0.100	0.016	0.126	0.128
$\beta_{Xt} = -0.02$	1.011	0.098	0.015	0.122	0.123
Cubic Polynomials					
$\beta_{Xt} = -0.2$	1.043	0.207	0.074	0.265	0.268
$\beta_{Xt} = -0.1$	1.018	0.121	0.023	0.148	0.152
$\beta_{Xt} = -0.04$	1.010	0.101	0.017	0.126	0.128
$\beta_{Xt} = -0.02$	1.016	0.099	0.016	0.122	0.124
B-Splines					
$\beta_{Xt} = -0.2$	1.039	0.206	0.073	0.263	0.267
$\beta_{Xt} = -0.1$	1.016	0.121	0.023	0.148	0.152
$\beta_{Xt} = -0.04$	1.008	0.101	0.016	0.126	0.128
$\beta_{Xt} = -0.02$	1.014	0.098	0.016	0.122	0.124
Automated Smoothing Splines					
$\beta_{Xt} = -0.2$	1.034	0.204	0.071	0.263	0.265
$\beta_{Xt} = -0.1$	1.012	0.120	0.023	0.147	0.151
$\beta_{Xt} = -0.04$	1.005	0.100	0.016	0.126	0.127
$\beta_{Xt} = -0.02$	1.011	0.097	0.015	0.122	0.123

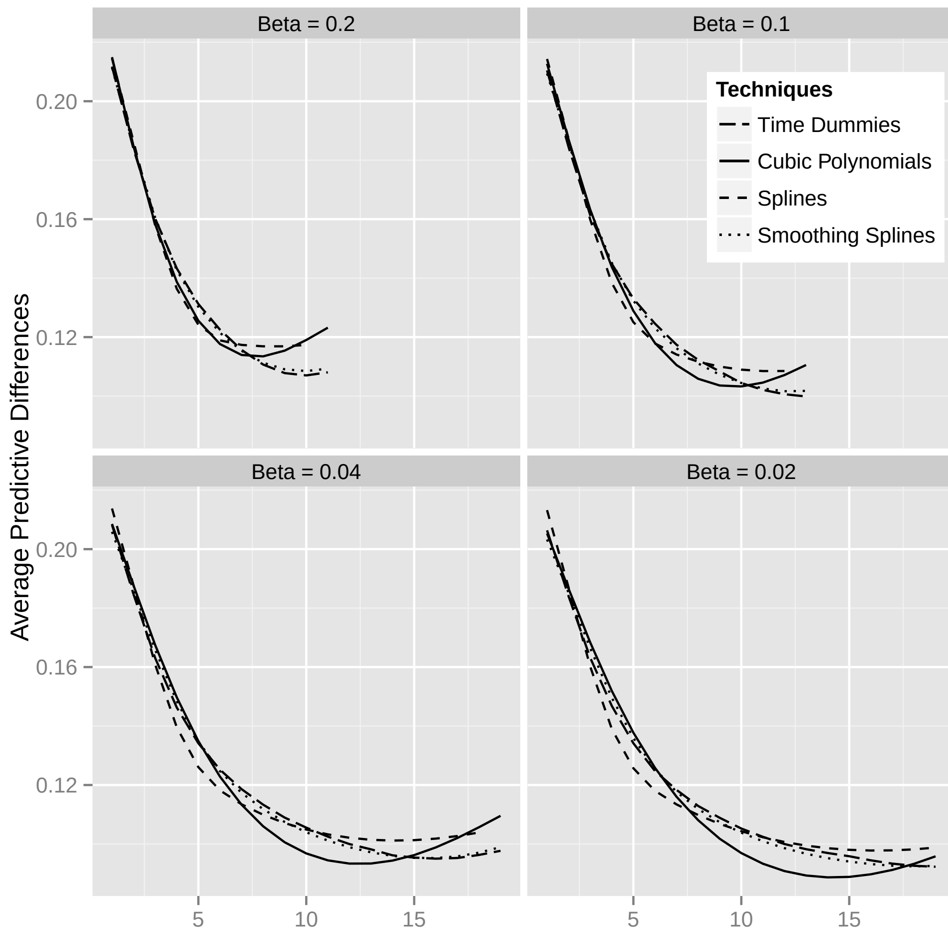
Note: N=1000, betas = c(-3,1), x is drawn randomly from a uniform distribution, $\in [-2, 2]$. **Bias** is the mean of absolute bias: $|\hat{\beta}_x - 1|$. **MSE** is the mean of expected squared bias: $E[(\hat{\beta}_x - 1)^2]$. **SE** is the mean of the simulated standard errors. **SD** is the standard deviation of the estimates.

Figure S.22: Average Hazard Rates Compared to True Hazard Rate under Non-Proportional Hazards



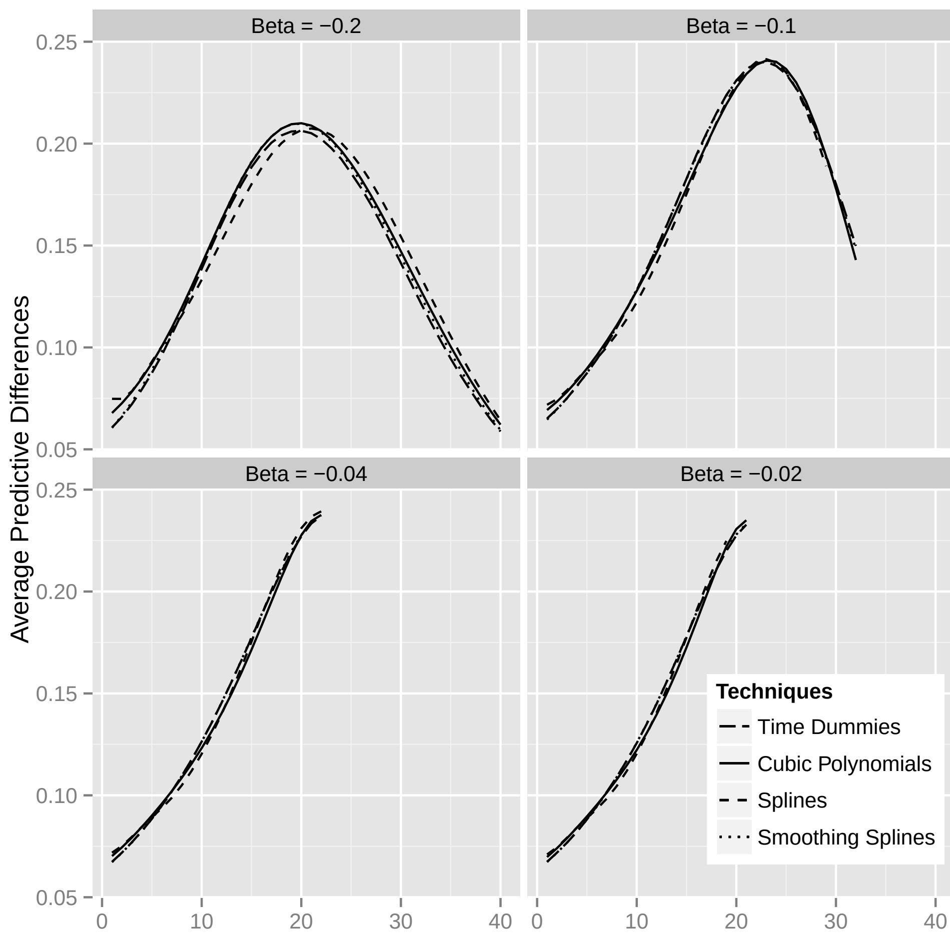
Note: N=1,000, sims = 1,000, beta = c(-3, 1), β_{Xt}=0.04 and -0.04, respectively.

Figure S.23: Average Predictive Differences of X_K ($\beta_{X_K} = 1$) with Alternative Estimation Techniques under Conditions Ignoring Non-Proportional Hazards



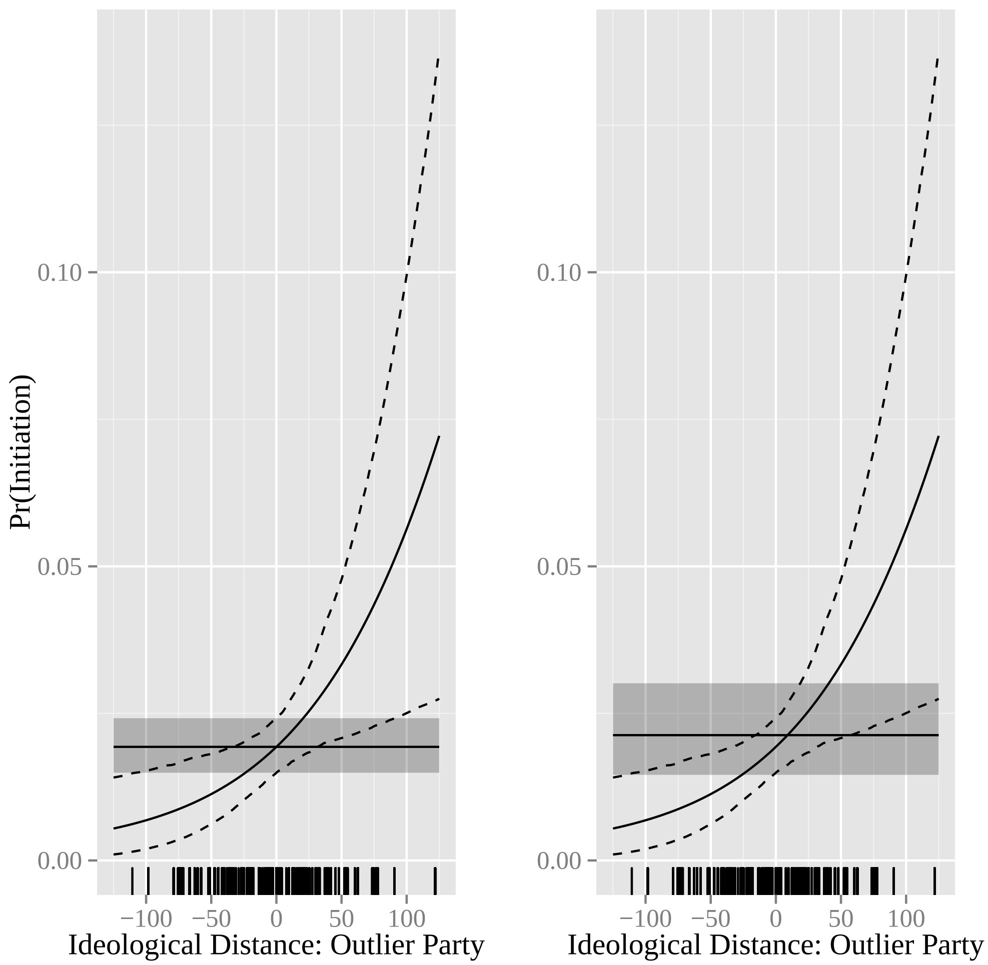
Note: The data reflect negative duration dependence with positive non-proportional hazards, $X_k \times t$. N=1,000, $\beta_0 = -3$. The non-proportional hazards are unmodeled in each technique.

Figure S.24: Average Predictive Differences of X_K ($\beta_{X_K} = 1$) with Alternative Estimation Techniques under Conditions Ignoring Non-Proportional Hazards



Note: The data reflect positive duration dependence with negative non-proportional hazards, $X_k \times t$. N=1,000, $\beta_0 = -3$. The non-proportional hazards are unmodeled in each technique.

Figure S.25: Predicted Probability of Dispute Initiation for Coalition Governments and *Ideological Distance* (Figure 2, Clare 2010: 982)



Note: Rug plots provide the sample distribution of *ideological distance*. Shaded areas represent the predicted probability of dispute initiation for cohesive coalitions (left panel) and single-party governments (right panel).

Figure S.26: Receiver-Operating Characteristic (ROC) Curve for Clare (2010)

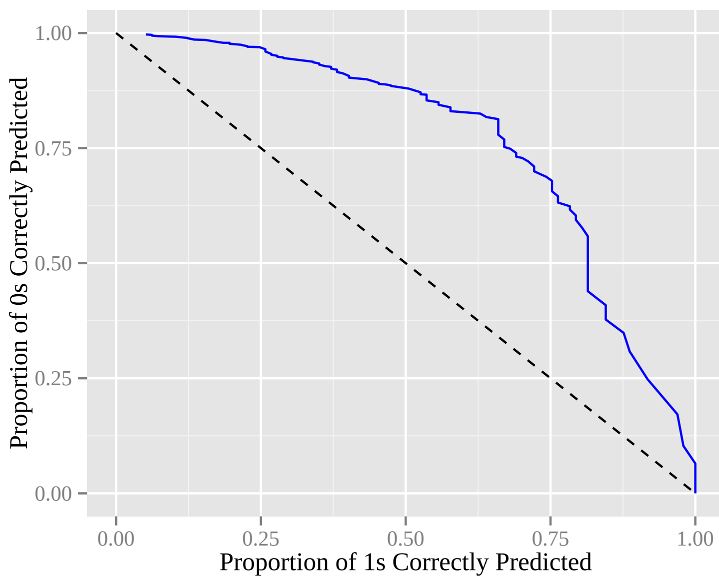


Figure S.27: Separation Plot for Clare (2010)

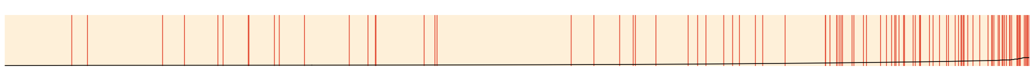


Table S.7: Values of Variables for Scenarios in Figure 6 of Manuscript

	Pr=0.003	**Pr=0.05**	**Pr=0.10**	**Pr=0.50**
Government Fractionalization	Min	75th	95th	Max
Peaceyears	5	5	5	5
Coalition Government	0	1	1	1
Government Ideology	Min	75th	80th	Max
Minority Government	1	0	0	0
Military Capabilities	25th	75th	90th	Max
Number of Bordering Democracies	75th	Mean	5th	Min
Number of Bordering Allies	75th	Mean	5th	Min
Trade Dependence	75th	Min	Min	Min
Number of Veto Players	75th	25th	5th	Min

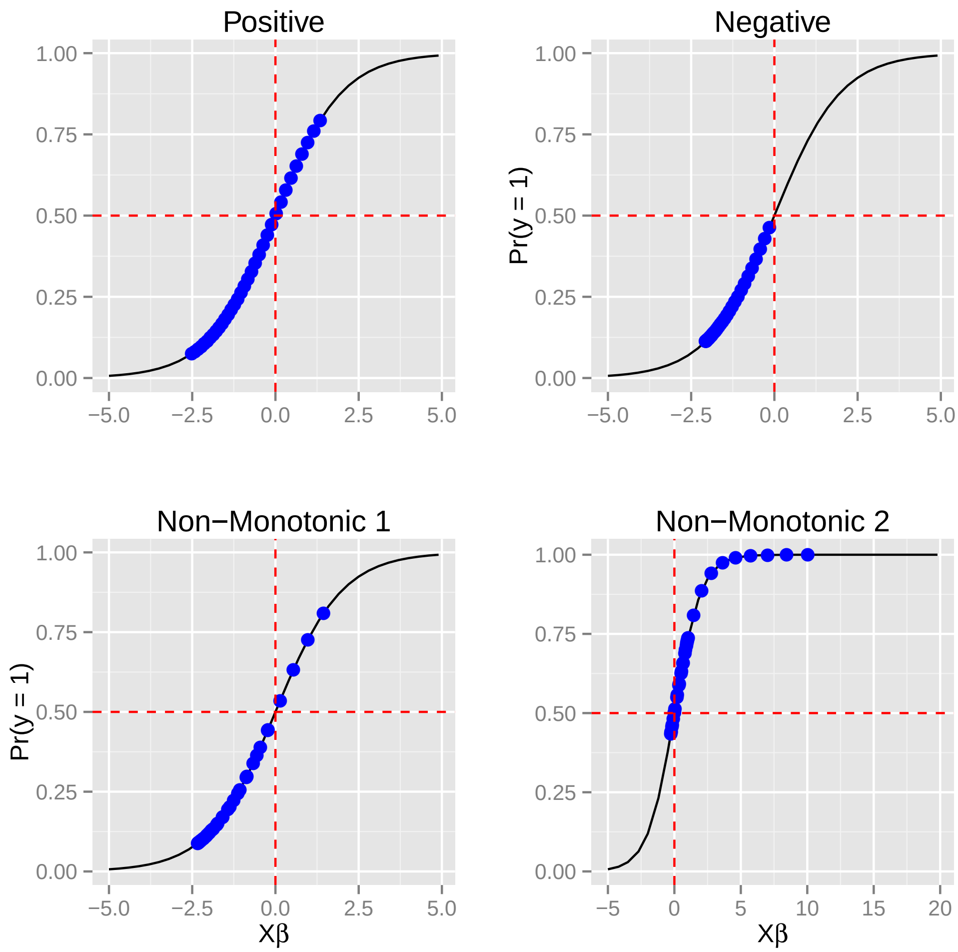
Note: Numbers represent either raw values (binary or discrete variables) or percentiles (continuous).

Table S.8: Test for Non-Proportional Hazards in Clare's (2010) Model of Dispute Initiation

Variable	β	S.E.
Government Fractionalization	0.007**	(0.003)
Peaceyears	-0.08**	(0.02)
Peaceyears^2	0.002**	(0.001)
Peaceyears^3	-0.00002**	(8.4×10^{-6})
Fractionalization$\times$Peaceyears	-0.0003	(0.001)
Fractionalization$\times$Peaceyears^2	2.2×10^{-6}	(4.0×10^{-5})
Fractionalization$\times$Peaceyears^3	2.9×10^{-9}	(4.6×10^{-7})
Coalition Government	0.04	(0.13)
Government Ideology	0.002	(0.001)
Minority Government	-0.23	(0.18)
Military Capabilities	5.10	(5.09)
Number of Bordering Democracies	0.14	(0.21)
Number of Bordering Allies	-0.30**	(0.11)
Trade Dependence	-0.06	(0.43)
Number of Veto Players	-0.05	(0.07)
N		3,336
ePCP		94.6%
AUC	0.77	0.03

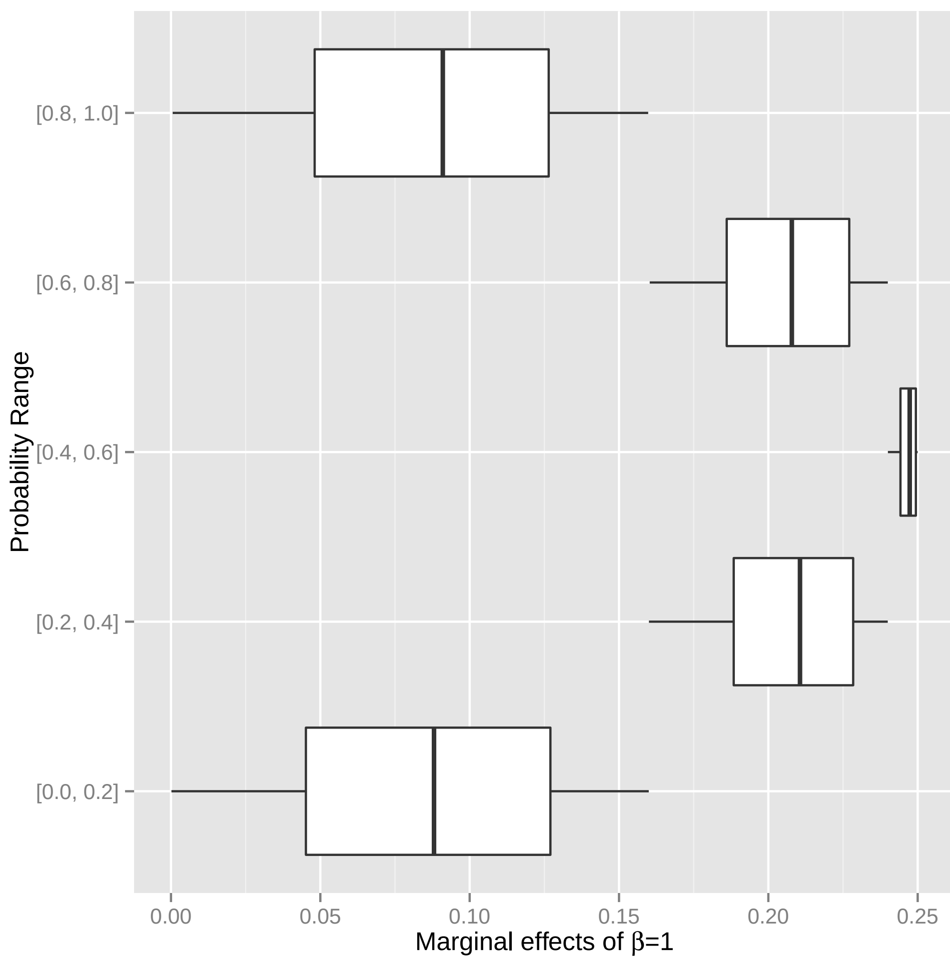
Note: $** = p < .05$, $* = p < .1$ (two-tailed)

Figure S.28: Varying the Values of the Temporal Dependence Variables Results in a Variety of Locations along the Logistic CDF: Four Functional Forms of Duration Dependence



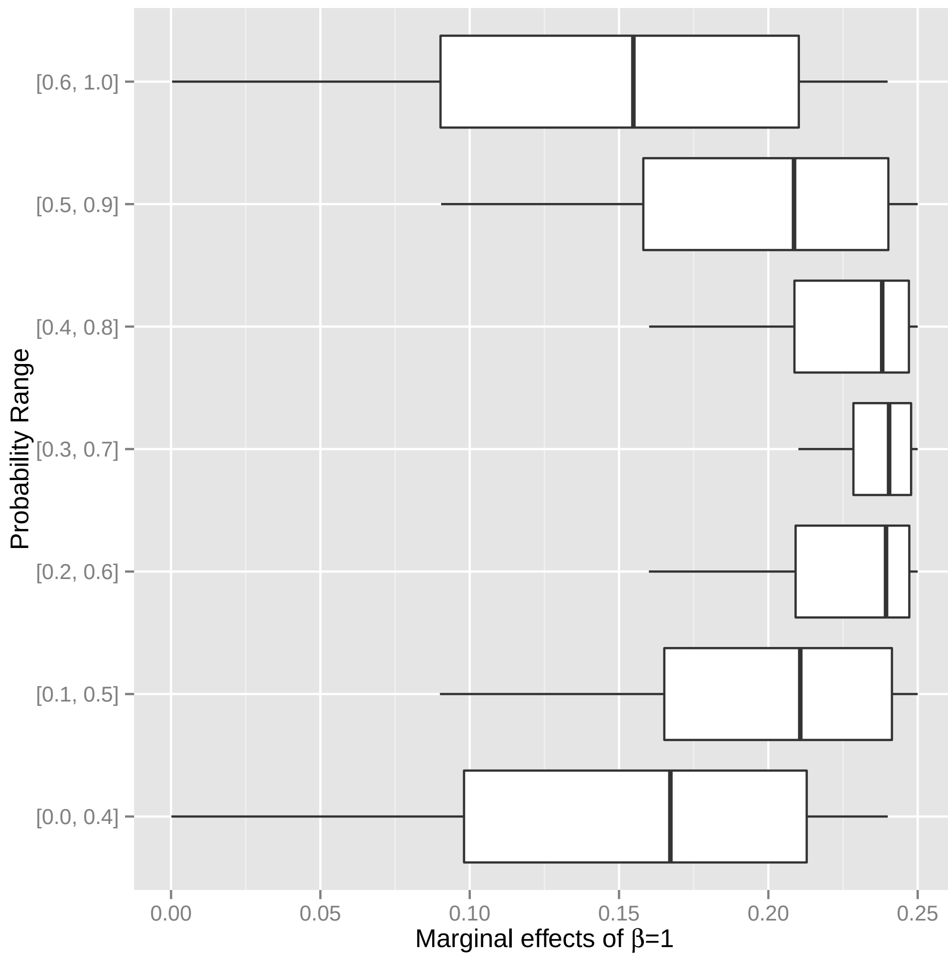
Note: Each dot represents the $X\beta$ value for the same observation, with varying values of the temporal dependence variables. As a reference, I include a solid line to represent the logistic CDF, and dashed lines to represent the middle of the CDF.

Figure S.29: Box-Whisker Plots of the Change in Predicted Probability for a 1-Unit Increase in x ($\beta = 1$) across Ranges of Probabilities: 0.2 Range



Note: Each range of probabilities is drawn from a uniform distribution ($N = 2000$).

Figure S.30: Box-Whisker Plots of the Change in Predicted Probability for a 1-Unit Increase in x ($\beta = 1$) across Ranges of Probabilities: 0.4 Range



Note: Each range of probabilities is drawn from a uniform distribution ($N = 2000$)