University of South Carolina
Scholar Commons

Faculty Publications | Political Science, Department of

Spring 2011

Change Comes with Time: Substantive Interpretation of Non-Proportional Hazards in Event History Analysis

Amanda A. Licht
University of South Carolina - Columbia, licht@mailbox.sc.edu

Follow this and additional works at: https://scholarcommons.sc.edu/poli_facpub

Part of the Political Science Commons

Publication Info

Postprint version. Published in *Political Analysis*, Volume 19, Issue 2, Spring 2011, pages 1-41.
This is a pre-copy-editing, author-produced PDF of an article accepted for publication in *Political Analysis* following peer review. The definitive publisher-authenticated version [Licht, A.A. (2011). Change Comes with Time: Substantive Interpretation of Non-Proportional Hazards in Event History Analysis. *Political Analysis*, *19*(2), 227-243. DOI: 10.1093/pan/mpq039] is available online at: http://pan.oxfordjournals.org/content/early/2011/03/29/pan.mpq039.

This Article is brought to you by the Political Science, Department of at Scholar Commons. It has been accepted for inclusion in Faculty Publications by an authorized administrator of Scholar Commons. For more information, please contact digres@mailbox.sc.edu.

Change Comes with Time: Substantive Interpretation of Non-Proportional Hazards in Event History Analysis

Forthcoming in *Political Analysis*
Spring 2011

Amanda A. Licht, PhD[1]
University of South Carolina

3rd Floor Gambrell Hall
Department of Political Science
Columbia, SC 29205

aalicht@gmail.com
http://people.cas.sc.edu/licht

Abstract: While methodologists have provided us ample notice of both the problem of non-proportional hazards and the means of correcting for them, less attention has been paid to the post-estimation interpretation. The suggested inclusion of time interactions in our models is more than a statistical fix: these corrections alter the substantive meaning and interpretation of results. Framing the issue as a specific case of multiplicative-interaction modeling, I provide detailed discussion of the problem of non-proportional hazards and present several appropriate means of interpreting both the substantive impact and the significance of variables whose effects may change over time.

[1] I owe any success this project may have to the gracious assistance of Andrew Enterline and Kyle Joyce, Scott Meinke, and Jonathan Golub and Bernard Steuenenberg for allowing me to replicate their studies. For close guidance and frequent reality checks, I must thank Fred Boehmke. Brian Lai, Mike Lewis-Beck, Sara Mitchell, and Jan Box-Steffensmeier all provided very helpful comments on early drafts. Thanks also are due to Rob Franzese, Mike Alvarez and anonymous reviewers for their assistance in improving the paper. Earlier drafts were presented at the summer meeting of the Society for Political Methodology poster session at New Haven in 2009, and at the American Political Science Association 2009 annual conference in Toronto.

1. Introduction:

In the last decade, Box-Steffensmeier and colleagues produced several explications of non-proportional hazards (NPH) as a serious statistical issue in event history analysis (Box-Steffensmeier and Zorn 2001; Box-Steffensmeier, Reiter and Zorn 2003; Box-Steffenesmeier and Jones 2004). As a result, models fitted for NPH became more common in the literature of all major subfields (e.g. Chiozza and Goemans 2004; Meinke 2005; Golub 2007; Golub and Steunenberg 2007; Balch-Lindsay, Enterline and Joyce 2008; Zhelyazkova and Torenvlied 2009). Correcting for NPH through the inclusion of time interactions for variables in violation of the proportional hazards assumption (PHA), however, is more than a quick statistical fix; it complicates the interpretation of statistical results and calls for more advanced post-estimation techniques. With the notable exception of Golub and Steunenberg (2007), little attention has been devoted to this aspect of NPH. Building on those authors' effort, I outline below three interpretation strategies which fit the bill. The simulation procedures I propose provide superior ability to report the substantive importance of continuous variables' NPH effects and appropriate information regarding uncertainty around those effects.[1]

Many analysts find interpreting the results of the NPH Cox model difficult. What does it mean, for instance, when the constitutive term's coefficient fits the hypothesis but the time interaction does not? Is the hypothesis supported, refuted or neither? What is the actual estimated effect of this variable, and how can we best convey it to the reader? Political scientists must answer these questions, because many of our hypotheses involve precisely the type of dynamic social processes likely to create NPH.

Acknowledging NPH corrections as a special case of multiplicative-interaction modeling, we can utilize existing knowledge to address these questions (see also Golub and Steunenberg 2007). As with any interaction effect, appropriate interpretation methods include the calculation of marginal effects and first differences (Brambor, Clark and Golder 2006; Kam and Franzese 2007). Extant software add-ons and guides, however, do not accommodate the complex, nonlinear Cox model.[2]

Discussion proceeds below with a brief review of the PHA and the proper interpretation of interaction effects. I then present the calculations for measures of substantive interest in the NPH Cox model and outline three methods for proper interpretation of non-proportional effects. Finally, I demonstrate the merits of these strategies through replication of NPH Cox models from three published papers covering civil war duration (Balch-Lindsay et al. 2008), directive approval in the European Union (Golub and Steuenenberg 2007), and U.S. Congresspersons' tendency to switch positions on recurrent issues (Meinke 2005). The broad substantive range of these studies illustrates the widespread need for attention to nonproportionally time-dependent processes in our expectations and our empirical models.

2. The Proportional Hazards Assumption and Political Science

The proportional hazards assumption (PHA) is embedded in the logic of the standard Cox duration model. To see this property consider the formula for its hazard:

$$h_i(t) = h_0(t)e^{\mathbf{X}_i\boldsymbol{\beta}}, \qquad \text{(Eq. 1)}$$

where i denotes each observation i= 1…n; t is a point in time; $\boldsymbol{\beta}$ is the vector of coefficient estimates and $\mathbf{X}_i$ is the vector of covariates. No matter the value of the linear

index, $\mathbf{X_i\beta}$, the form of the function dictates that risk of failure at each point in time shifts "proportionally" by the baseline, h_0. This issue may be more clearly visible in the hazard ratio function, which expresses the risks of failure for a given case i relative to that of another case j.

$$h_i(t)/h_j(t) = e^{\boldsymbol{\beta}'(\mathbf{X_i}-\mathbf{X_j})}. \qquad \text{(Eq. 2)}$$

The effect of covariates X may shift from one unit of time to the next (e.g. from day one to day two, or from day ten to eleven), but always by this single, proportional quantity.

Box-Steffensmeier and Jones (2004, 132) warn: "Theoretically, one would be suspicious of the [PHA] if there was reason to expect that the effect of a covariate changes over time".[3] As political scientists, our suspicions regarding the PHA should be aroused more often than not. While the term "non-proportional hazards" sounds esoteric, situations giving rise to it abound. If we argue that experience makes people better at their jobs, that the effectiveness, tenability and legitimacy of institutions increases with age, that actors in competition send signals and adjust their strategies in response to each other, that new ideas gather force from their transmission until reaching an important threshold, then we are arguing that key covariates' effects will change over time. If these dynamic trends do not modify the effects of all covariates equiproportionally over time, then the proportional hazards assumption will not hold. Essentially, the nature of the political processes of learning, institutionalization, strategic developments and information transmission which interest us as a discipline are likely to produce frequent violations of the PHA.

Finnemore and Sikkink's (1998) theory of international norm life-cycles, for example, hypothesizes a predictable, increasing trend in the power of norms to alter

actors' behavior as they become internalized over time. Models of norm-efficacy typically pit the importance of ideas against that of power and strategic interests. While norms strengthen over time, geopolitics and physical power will exert either static or decreasing pressure on international actors. Another widely known theory posits operation of a "generational effect" on the repugnance for war within a society; the grandchildren of soldiers tend to romanticize warfare, having suffered none of its costs directly (Wright 1942; Toynbee 1972). This would produce a pattern of decreasing likelihood of conflict following a traumatic war, the magnitude of which would decay over time until the opposite effect would ultimately surface with the growing influence of hawkish grandchildren. While the worldview of the citizens follows this dynamic path, the strongest determinants of international war – territory and rivalry – will likely retain effects of the same direction and magnitude over time.

Theories which imply effects likely to violate the PHA also exist in comparative and American politics. Light's (1999) analysis of the American presidency posits two competing, simultaneous dynamic processes. The first is an increasing trend in presidential competence over time: all presidents learn on the job and all presidents try to improve over time. The second is a decreasing stock of political capital: as time passes, presidential approval inevitably declines as opposition criticism increases. Unless these dynamics shifted proportionally with the effects of all other covariates included, they would certainly violate any event-history analysis of a president's efficiency and effectiveness assuming proportional hazards.[4] The opposing direction of these dynamics likely increases the odds that any model attempting to account for both will violate the PHA.

In comparative politics, the study of electoral institutions' effects on voter behavior and electoral outcomes ranks among the oldest and most prolific research agendas (e.g. Cox 1997; Duverger 1954). The equilibrium behaviors suggested by different electoral institutions, however, cannot be expected to materialize instantly. Voters require time and experience with a new system to determine how to vote "strategically"; the effect of institutions on behavior, then, will change over time as voters observe the results of their actions (e.g. Wittrock 2008). Theories of political party emergence and competition also posit dynamic learning and institutionalization processes which would produce changing effects over time (e.g. Kitschelt 1989; Downs 1997). Inglehart's modernization thesis predicts a changing effect of economic development and democratic government on a host of questions over the course of time (e.g. Inglehart and Welzel 2005). In all three cases, these institutional changes take place in a context of perhaps stickier cultural and geographical effects. Individual-level explanations of voter-behavior, also, must contend with the generally more static effects of partisan identification, race and gender. If these theories hold, then we should expect the effects of some variables in event-history models of voting behavior to change over time. Because of the mixture of social processes and structural effects, it seems unlikely that all these effects will shift proportionally over time.

The examples above outline reasons to believe the PHA will be violated in many analyses within all major branches of political science research. When the PHA is violated without correction, systematic bias in coefficient estimates results: the effects of variables with NPH are overestimated while the hazard rate increases and underestimated while it decreases. Moreover, if the effects of such variables change from positive to

negative over time (or vice versa) the violation of the PHA may result in a Type II error, declaring the effects insignificant when they in fact have a conditionally significant, but changing, effect (Box-Steffensmeier and Jones 2004, 132). Statisticians have developed strategies for assessing the violation of the PHA in the Cox model, notably including the schoenfeld residuals test (Box-Steffensmeier and Zorn 2002, 977). This test is easily implemented in statistical packages.

Box-Steffensmeier and colleagues recommend a simple response to significant results of the Schoenfeld residuals test: model the relationship between covariates and time by including interaction terms, relating the two by some standard function (Box-Steffensmeier and Jones 2004, 136), most commonly the natural log of time (Box-Steffensmeier and Zorn 2002, 978; Box-Steffensmeier and Jones 2004, 136; Box-Steffensmeier et al. 2003). Inclusion of these multiplicative-interaction terms alongside the constitutive term allows estimation of both the initial effect and the over-time changes in effect.[5] The proliferation of interactions, however, substantially complicates interpretation of results.

Of course, this is no reason to eschew appropriate model specification. The analyst must take care to assess, anticipate, and evaluate deviations from the PHA and model these trends appropriately, recognizing that inclusion of multiplicative-interaction terms alters the meaning of the estimated coefficients and standard errors. Since Friedrich's (1982) essay on the subject, many other social scientists have written in defense of multiplicative modeling and detailed the issues it raises regarding interpretation and reporting uncertainty (see Jaccard and Turrisi 2003; Brambor et al.

2006; Kam and Franzese 2007). The current wisdom on the interpretation of multiplicative-interactions will be briefly reviewed in the next section.

3. The Interpretation of Multiplicative Models

As Kam and Franzese (2007, 20) bluntly pronounce, "... *coefficients* [on constitutive and interaction terms] *are not effects*"; they are not as readily interpretable as in purely additive-separable linear regression. The constitutive terms' coefficients tell us the effect of that variable when the other interacting variable(s) is (are) equal to zero. Accordingly, "… rather than being constant (as they are in the additive model), the standard errors of the conditional coefficients vary according to the level of the other independent variable" (Friedrich 1982, 810).[6] With the proliferation of non-linear modeling techniques, quantitative political scientists have become accustomed to performing minor transformations, such as retrieving the odds-ratio or calculating predicted probabilities, to retrieve substantively meaningful quantities from their regression coefficients. Inclusion of interaction effects introduces a similar need for additional post-estimation calculations (Friedrich 1982; Ai and Norton 2003; Brambor et al. 2006; Kam and Franzese 2007). Obviously, we would like to know about our independent variables' impact on the dependent variable at more than just one (possibly rare, non-existent, or even illogical) value of the other factor. Political methodologists forward two useful strategies for obtaining such information: marginal effects and first differences (see Brambor et al. 2006; Kam and Franzese 2007).

The marginal effect of a covariate, X_1, which has been interacted with another covariate, X_2, is the first derivative of the likelihood function, $\partial Y/\partial X_1$. In the linear

regression case, the marginal impact of an increase in X at a specific point is simply the sum of the effect of X_1 when X_2 is zero (i.e. the constitutive term coefficient) and the the product of the interaction term coefficient and the value of X_2 (see Friedrich 1982; Brambor et al. 2006; Kam and Franzese 2007).[7] $\partial Y/\partial X_1$ is, by definition, the change in the dependent variable for a marginal change in the independent variable. In other words, marginal effects tell us the change in Y for an arbitrarily small change in X_1.

The information necessary for calculation of marginal effects and their standard errors for most models is easily attained using modern statistical software. Brambor et al. (2006) provide template code for the necessary procedure for linear regression and binary dependent variable models, which should allow even inexperienced programmers to obtain appropriate results and present them graphically.[8] The `grinter` utility, written for Stata by Boehmke, allows analysts to produce equivalent graphics with even less effort.[9]

When X_1 is an ordinal or nominal variable, however, the marginal effect approach can be an unduly abstract means of communicating results. Kam and Franzese (2007, 25) and King et al. (2000) thus recommend differences in predicted probability as an alternative. A first differences approach calculates the change in the dependent variable as the independent variable moves from one discrete value to another, taking bigger steps than the marginal changes underlying differentiation. The first difference may be calculated for unit changes or for substantively interesting intervals (e.g. from the minimum to the mean). Standard errors and first differences from many models can be calculated using the guidelines available in Kam and Franzese (2007, 89) or with King et al.'s (2000) Clarify software for Stata.

4. Interpretation Strategies for the NPH Cox Model

Calculating appropriate marginal effects and standard errors for the NPH Cox model may be daunting for many political scientists, due to the complex, nonlinear functions involved. Under conditions of nonproportional hazards, an estimate of the Cox hazard rate for case j at time t is,

$$h_j(t) = e^{(\beta_1 X_j + \beta_2 \ln(t_j) X_j + \boldsymbol{\beta Z_j})} h_0(t). \qquad \text{(Eq. 3)}$$

The exponential term gives the sensitivity of the hazard to covariates, X and **Z**; unlike in the standard PHA model, the non-proportionally time-dependent effects of X are modeled by the interaction of X with log time. The baseline hazard rate, $h_0(t)$ is the probability that a case will fail at time t, a value which is assigned equally to all cases j which are still "at risk" (i.e. which have not yet failed). To recover this quantity, we must calculate the likelihood of a case j remaining in the risk set, R, at time $t=t_i$. That likelihood is then subtracted from one to yield the baseline hazard for the time threshold $t=t_i$:

$$\hat{h}_0(t_i) = 1 - \sum_{j \epsilon R(t_i)} e^{(\beta_1 X_j + \beta_2 \ln(t_i) X_j + \boldsymbol{\beta Z_j})}. \qquad \text{(Eq. 4)}$$ [10]

Even in this simplest possible scenario which includes only one PHA-violating effect and a single level interaction, taking the derivative with respect to X produces a complex function.[11] As the number of PHA-violating effects proliferates, the difficulty of calculating this function and its standard error will increase. Happily, those who wish to avoid the time and effort necessary to calculate the true marginal effect may do so by following Golub and Steunenberg's (2007) example.

Golub and Steunenberg (2007) utilize a helpful algebraic manipulation. By constructing a ratio of the hazards of two hypothetical cases i and j that differ only with respect to the value of the covariate X, we can define, at time t=t,

$$h_i(t)/h_j(t) = \frac{h_0(t)e^{(\beta_1 X_i + \beta_2 \ln(t) X_i + \boldsymbol{\beta Z})}}{h_0(t)e^{(\beta_1 X_j + \beta_2 \ln(t) X_j + \boldsymbol{\beta Z})}} = 1\left(\frac{e^{(\beta_1 X_i + \beta_2 \ln(t) X_i + \boldsymbol{\beta Z})}}{e^{(\beta_1 X_j + \beta_2 \ln(t) X_j + \boldsymbol{\beta Z})}}\right)$$

$$= e^{(\beta_1 X_i + \beta_2 \ln(t) X_i + \boldsymbol{\beta Z})} e^{-(\beta_1 X_j + \beta_2 \ln(t) X_j + \boldsymbol{\beta Z})}$$

$$= e^{(\beta_1 X_i + \beta_2 \ln(t) X_i + \boldsymbol{\beta Z}) - (\beta_1 X_j + \ln(t) X_j + \boldsymbol{\beta Z})}$$

$$= e^{(\beta_1 X_i - \beta_1 X_j) + (\beta_2 \ln(t) X_i - \beta_2 \ln(t) X_j)}$$

$$= e^{(X_i - X_j)(\beta_1 + \beta_2 \ln(t))} \qquad \text{(Eq. 5).}$$ [12]

This hazard ratio transformation removes the ungainly mathematical issues raised by the baseline hazard, thus dramatically reducing the difficulty of the calculations needed to interpret the role of variables whose effects change over time.

Golub and Steunenberg (2007) utilize two further simplified measures. First, they discuss a special case of the hazard ratio, constructed such that the covariate X in the hypothetical observation j equals zero. Under this condition, the hazard ratio simplifies:

$$h_i(t)/h_j(t) = e^{X_i(\beta_1 + \beta_2 \ln(t))} \qquad \text{(Eq. 6).}$$

The authors refer to this quantity as the relative hazard. When X is binary, it describes the change in hazard when the variable in question is "switched on". When X is continuous, it gives the change in the hazard resultant from a jump to $X=x_i \neq 0$ from $X=0$.

Secondly, Golub and Steunenberg (2007) examine the combined coefficient, $(\beta_1 + \beta_2 \ln(t))$,[13] which relates to the contribution of X_i to the hazard rate (via the model's exponential-function transformation) over the range of time Analysts can easily obtain accurate standard errors for this quantity, because the well-known formula for the variance of a sum of random variables used in the linear-interaction context applies (see Friedrich 1982, Brambor et al. 2006, Kam and Franzese 2007). However, reporting uncertainty regarding the relative hazard in Equation 6 is nearly as simple.

[FIGURE 1 HERE]

Exponentation creates the relative hazard as a monotonically increasing function of the combined coefficient. While the variance of the relative hazard will be greater than the variance of the combined coefficient, its distribution will correspond to that of its referent. Figure 1 charts a relative hazard against its non-exponentiated form.[14] The reference lines indicate key points in the distribution of both variables, the fifth, fiftieth and ninety-fifth percentiles. The 90% confidence interval around the relative hazard can be recovered by exponentiating the corresponding confidence bounds around the combined coefficient.[15]

The combined coefficient and relative hazards approaches provide viable and easily implemented means of evaluating the effects of time-dependent variables. They allow evaluation of statistical significance across the full range of time and straightforward reporting of estimation uncertainty. In terms of relating the substantive importance of an effect, however, these methods, as employed by Golub and Steunenberg (2007), may fall short. The combined coefficient describes the direction and magnitude of the variable's effect on the hazard rate over time at a specific value (usually x_i=1) held constant over time. While interesting, this quantity is abstract. It may not tell us whether the variable exerts a substantively notable impact and it cannot reflect movement in the variable over time.

The relative hazard (see Eq 6) offers a more substantive interpretation for binary NPH effects, describing the difference in hazard given an X-unit change in the variable of interest. With a continuous X, however, this measure may have less substantive appeal. The relative hazard requires a shift to some value $X=x_i$, from x_j=0, a value which may be

meaningless or impossible. The effect of a continuous, X which has been interacted with a function of time depends on the specific value x_i which is being used as well as the value of time. A single calculation of the relative hazard, or combined coefficient will not suffice to describe the full content. Consistent with the general good practice of utilizing either first differences or marginal effects to interpret continuous-variable interaction effects, I propose two simple simulation techniques appropriate for evaluating continuous NPH effects.

The first method for dealing with non-binary, time-dependent effects returns to the hazard ratio expressed in Equation 5. The hazard ratio can be utilized in this previously presented form, or, as recommended by Box-Steffensmeier and Jones (2004, 60), transformed into a percentage change in hazard rate for added substantive appeal. Adopting the notation in use here, the percentage change in hazard rate at time t=t can be calculated as:

$$\%\Delta h_i(t) = \left(\frac{h_0(t)e^{(\beta_1 X_j + \beta_2 \ln(t) X_j + \boldsymbol{\beta Z})} - h_0(t)e^{(\beta_1 X_i + \beta_2 \ln(t) X_i + \boldsymbol{\beta Z})}}{h_0(t)e^{(\beta_1 X_i + \beta_2 \ln(t) X_i + \boldsymbol{\beta Z})}}\right) \times 100 \qquad \text{(Eq. 7).}$$

This formula simplifies:

$$\%\Delta h_i(t) = \left(\frac{e^{(\beta_1 X_j + \beta_2 \ln(t) X_j + \boldsymbol{\beta Z})}}{e^{(\beta_1 X_i + \beta_2 \ln(t) X_i + \boldsymbol{\beta Z})}} - \frac{e^{(\beta_1 X_i + \beta_2 \ln(t) X_i + \boldsymbol{\beta Z})}}{e^{(\beta_1 X_i + \beta_2 \ln(t) X_i + \boldsymbol{\beta Z})}}\right) \times 100$$

$$= \left(e^{(\beta_1 X_j + \beta_2 \ln(t) X_j + \boldsymbol{\beta Z}) - (\beta_1 X_i + \beta_2 \ln(t) X_i + \boldsymbol{\beta Z})} - 1\right) \times 100$$

$$= \left(e^{(X_i - X_j)(\beta_1 + \beta_2 \ln(t))} - 1\right) \times 100 \qquad \text{(Eq. 8).}$$

As this measure closely resembles a traditional first difference in interpretation, I will refer to it as the first differences method rather than as a change in hazard ratio. For a given change in X, the effect is null when Eq. 8 is statistically indistinguishable from zero, rather than from one. X_i and X_j can be selected to provide a good measure of

substantive significance. When the effect of X is suspected to change substantially across its range of values, it may be wise to calculate a first difference at multiple values of X_i and X_j. The replication of Meinke's (2005) vote-switching study below provides an example of this situation.

A simple simulation process can ease the calculation of confidence intervals around either the hazard ratio or the first difference. As in King et al.'s (2000) Clarify software, a large number of $\widehat{\boldsymbol{\beta}}$ can be drawn from a multivariate Normal distribution with mean and variance taken from the parameter and parameter-covariance estimates of the NPH Cox model. This procedure approximates the sampling distribution of $\widehat{\boldsymbol{\beta}}$, allowing us to present the best estimates complete with confidence bounds at the desired level as follows:

1. Create a dataset with T observations, and generate values of time $T=[t_1 \ldots t_T]$;[16]
2. Draw vector of $\widehat{\boldsymbol{\beta}}$ using the parameter and covariance estimates;
3. Calculate the desired first difference (Equation 8) or hazard ratio (Equation 5);
4. Repeat steps 1-3 N times;
5. Preserve median and percentile values appropriate to desired confidence level for each $T=t_i$ and (X_i-X_j).[17]

A second strategy for interpreting a continuous, PHA-violating effect calculates the relative hazard (Equation 6) at many values of X and across the range of time. This can be accomplished with the following steps:

1. Create a dataset with T observations, and generate values of time $T=[t_1 \ldots t_T]$;
2. Draw a vector of $\widehat{\boldsymbol{\beta}}$ using the parameter and covariance estimates;
3. Calculate the relative hazard (Equation 6) at value x_i of X;

4. Repeat steps 1-3 N times;
5. Repeat steps 1-4 at some, many, or all interesting values of X, compiling results in one dataset;
6. Preserve median and percentiles appropriate to desired confidence level for each $T=t_i$ and $X=x_i$.

These simulation procedures will generate the median and upper and lower percentiles of the "confidence interval" for the measure at each combination of time and covariate values. The measure and confidence bounds can then be charted against time to present a graphical summary of the effects' substantive and statistical significance. While a few studies have utilized these or similar means of interpretation (e.g. Box-Steffensmeier and Zorn 2001; Golub and Steunenberg 2007; Zhelyazkova and Torenvlied 2009), the appropriate processes have not been elaborated transparently, which may have abetted the continued confusion across most empirical, political science research. In the next sections I hope to dispel this remaining confusion, first by discussing selection of a proper measure for interpreting NPH Cox models, and providing tips for hypothesis testing. The following replication of three empirical studies will provide further details and examples.

5. Substantive Interpretation of Non-Proportional Effects in the Cox Model

Interpreting PHA-violating effects requires fulfillment of four tasks: (1) calculation of the variable's effect, (2) evaluation of that effect over time, (3) substantive assessment of that effect, and (4) incorporation of uncertainty. Above, I recommended three measures of substantive significance which can fulfill these criteria: the hazard

ratio (Equation 5), the first difference (Equation 8), and the relative hazard (Equation 6).[18]

Though each measure brings helpful attributes to interpretation, the differences between them are cosmetic. Subtracting one and multiplying by 100 changes the relative hazard in Equation 6 into the $\%\Delta h_i(t)$ in Equation 8 (if X_i-X_j=1). Similarly, the hazard ratio in Equation 5 and the relative hazard differ simply by the value assigned to X_j: when X_j=0, the relative hazard and hazard ratio are equal. If the change in X remains equal, identical conclusions will be drawn regarding statistical significance.

This mathematical equivalence allows the analyst to choose the most appropriate tool for his/her own presentational purposes. The two simulation procedures outlined above work well for continuous variables, but would be unnecessary for a PHA-violating effect which stems from a dichotomous variable. Because a change of less than one unit makes little sense for a binary variable, the hazard ratio simplifies to the relative hazard. Using simulation procedures to calculate the confidence intervals would be unnecessarily time-consuming.[19]

For continuous PHA-violating effects, choosing between the hazard ratio (Equation 5) and the first difference (Equation 8) will be driven more by stylistic preference and intuitive appeal than by efficiency. For example, in studies of international and civil conflict, a hazard ratio could report a raw change in the hazard of war onset as a tiny number. Given the low probability of interstate war, however, the relative change may be huge. In such cases, the analyst may believe a percentage change more accurately reflects substantive importance.[20]

Before embarking on the full process of calculating substantive measures, one may begin interpreting NPH Cox results through the returned coefficients. The constitutive coefficient relates to the effect of the variable when the log of time equals zero (i.e. when the time counter equals one). As most hypotheses do not specifically refer to t=1, and since the scaling of the time counter is arbitrary,[21] analysts must take care: a correctly signed constitutive coefficient need not translate to hypothesis support. The fate of our hypotheses rests largely with the overtime trend. When the interaction term holds the same sign as the constitutive coefficient, the initial effect magnifies over time. Commonly, however, the trend opposes the constitutive effect. Though often interpreted as "decay" in the original effect, this need not imply that the effect of a variable is "wearing off" over time (Golub and Steuenenberg 2007). The overtime trend may swamp the initial effect, both statistically and substantively.

The relative size and sign of the interaction and constitutive terms' coefficients indicate the rate at which X's effect moves over time. A small (large) interaction term coefficient relative to the constitutive term coefficient indicates slow (fast) change. When the coefficients have opposing signs, exponentiating the ratio of the coefficients provides the value of time when the estimated effect flips from positive to negative or vice versa (Box-Steffensmeier and Zorn 1998, 17). More precisely, that value, T_f, is:

$$T_f = e^{\frac{|\beta_1|}{|\beta_2|}} \qquad \text{(Eq. 9).}$$

Locating T_f on the distribution of failure times allows evaluation of the proportion of cases experiencing estimated effects in either direction. Chiozza and Goemans (2004, 610), for example, use this method to demonstrate that the institutional instability of mixed regimes heightens leaders' risks of losing office for the first 17 years, at which

point the estimated effect changes sign. Because 17 years is an extremely long time to hang onto office, the destabilizing effect applies to most mixed-regime leaders. Here T_f reveals some interesting, hypothesis-relevant information.[22] Evaluating coefficients, then, can be helpful, but these strategies tell us nothing about how the variables' effects evolve over time. The replications below illustrate the added benefit of utilizing the approaches discussed in Section 4.

5.1 Replication: Golub and Steunenberg (2007)

Golub and Steunenberg (2007) examine the duration of European Union (EU) deliberation on directives. Findings related to two key independent variables will be reviewed here: qualified majority voting (QMV) and legislative backlog.[23] Table 1 displays the successful replication of Golub's original (2007) analysis.

[TABLE 1 HERE]

Figure 2 demonstrates the relationship between the relative hazard and first-differences approaches for binary NPH effects. The left-hand panel replicates Golub and Steunenberg's figure (2007, 561), charting the relative hazard of QMV pre- and post-Single European Act across deliberation time. The right panel displays the first-differences strategy. Comparing across the Y-axis confirms the first difference curves differ from the relative hazard curves by exactly one unit.[24] Substantively, proposals not requiring unanimity clear faster, with QMV-eligible legislation initially 300% more likely to be settled than with unanimity required. The voting rule matters most in the early stages of deliberation; contentious proposals prolong the process, regardless of rules.

[FIGURE 2 HERE]

To interpret the effect of backlogged legislation, Golub and Steunenberg (2007, 560) calculate its combined coefficient at several points in time, holding the backlog itself constant at one.[25] Using this technique, the authors conclude that the very small effect of legislative backlogs achieves statistical but not substantive significance.

[FIGURE 3 HERE]

Unfortunately, this strategy is somewhat misleading regarding the substance of the estimated effect. The combined coefficient relates only the effect of one backlogged item, an unlikely situation for any regularly operating legislative body. The EU backlog averaged 169 pieces during the observation time. To test the backlog's impact on deliberation speed at more realistic values, I used Equation 6 and the simulation process above to chart its relative hazard from the minimum to the maximum. This strategy produced Figure 3, in which shaded bars demarcate the range between the 2.5th and 97.5th percentiles of estimated relative hazards. Lighter shading indicates larger backlogs. The thin black line and hollow diamonds highlight the median estimated relative hazard of one backlogged item, which represents Golub and Steunenberg's (2007, 560) strategy. I split the figure to ease evaluation of the two periods of significant effects.

Backlog impact increases as legislation accumulates. Compared to the hypothetical effect which Golub and Steunenberg (2007) evaluated, even the smallest observed backlog of less than 40 pieces of legislation exerts a powerful initial acceleration in deliberation. At this value, the likelihood of proposals being handled in the first days is 50% higher than if there were no backlog. When the backlog approaches the mean, the relative hazard nears a 500% increase. Over time, the escalatory effect of

the backlog declines, becoming statistically insignificant for a three year range (including the mean and modal durations).[26] For more controversial pieces of legislation, however, average or larger backlogs may further prolong deliberations. Combining an average backlog with a long deliberation produces a likelihood of settlement less than half that of a similar bill given no backlog.[27] Such a significant negative impact suggests that a large queue of work not only fails to "… expedite passage of the most controversial pieces of legislation" (Golub and Steunenberg 2007, 564), but can actually retard those processes.

5.2 Replication: Balch-Lindsay, Enterline and Joyce (2008)

Balch-Lindsay et al. (2008) study the effect of external intervention on the likelihood of civil war termination via government victory, opposition victory or negotiated settlement. Table 2 presents the successfully replicated results of their NPH Cox model with competing risks. Discussion below focuses on the key factor in the original analysis – intervention on behalf of the government.

[TABLE 2 HERE]

This replication provides an example of a common question in NPH Cox evaluation: how to evaluate hypotheses when constituent coefficients and time interactions defy expectations. Balch-Lindsay et al. (2008) predict a positive relationship between intervention and government victory, and a negative relationship between intervention and negotiated settlements. The estimated constitutive term coefficients, however, are positive and significant in both columns: intervention in favor of the government increases the likelihood of both outcomes at time ln(t)=0, which is, here, the

first day of a civil war. The negative, significant interaction terms suggest a changing effect, leading us to the overtime trend.

Figure 4 charts the combined coefficients for pro-government intervention over time.[28] The thick lines indicate the combined coefficient for pro-government intervention in each model, with thin lines marking the 90% confidence intervals.[29] The dashed curves depict the density of civil war duration times.

[FIGURE 4 HERE]

The left-hand panel displays intervention's effect on the likelihood of government victory. The rate of change in effect quickly drags the initially positive estimated effect into negative territory. This negative effect becomes significant at the 90% level at day 286 of civil war duration. Only about 15% of in-sample civil wars terminate before this date, suggesting that intervention negatively impacts the likelihood of government victory in most civil wars. Exponentiating the combined coefficient at day 286 reveals a relative hazard of $e^{-1.2}$=.30; civil wars with intervention experience a 70% reduction in the likelihood of government victory. The negative effect continues to grow over time, topping out at an estimated 95% reduction at the longest observed duration time.

The right-hand panel of Figure 4 displays the effect of intervention on the likelihood of negotiated settlements. Contrary to the authors' prediction, the effect never achieves significance. At the 90% threshold, its confidence bounds include zero for the entire duration of the longest observation time. Charting the full effect of this time-dependent variable in both equations, then, reveals findings contradictory to expectations.

[FIGURE 5 HERE]

Figure 5 replicates the first reported figure from Balch-Lindsay et al. (2008, 358), allowing comparison of interpretation techniques. Figure 5 plots the predicted survival probability of each competing outcome with pro-government intervention turned on. The authors note that the figure illustrates "… a third-party intervention on the side of the government has a different effect on the expected probability of a civil war surviving past a given point in time for each civil war outcome" (Balch-Lindsay et al. 2008, 356). The graph does suggest a much higher probability of government victory or negotiated settlement than of opposition victory. This apparent effect, however, cannot be safely attributed to the pro-government intervention. Plotting the coefficient, as in Figure 4, reveals that any increase in the chances of these outcomes could or should be attributed to some other shared characteristic: the only significant effect which intervention exerts on these termination types is negative.

5.3 Replication: Meinke (2005)

The dynamic nature of Congressional politics in Meinke's (2005) analysis turns our attention to the NPH model's benefits for political science. Meinke (2005) specifies a NPH Cox model of vote-switches regarding the minimum wage. Table 3 contains the successfully replicated results.[30] We will re-evaluate two variables: an indicator for losing the executive and a continuous measure of electoral security (the incumbent's mean-centered share of the vote).

[TABLE 3 HERE]

[FIGURE 6 HERE]

For the binary PHA-violating effect of losing the White House, charting the relative hazard, as in Figure 6, displays both statistical and substantive significance. For freshman congresspersons, a change in presidential party coincides with an estimated 100% increase in risk of switching. After this year, the effect of a change in executive control diminishes by half. Within three years, it loses statistical significance.[31] This precipitous decline could result from establishment of visible voting records, or from strategy-shifts towards constituent relations rather than partylines.

The time-varying effect of electoral security illustrates the value of fully charting continuous NPH effects. Though the coefficient on the interaction term falls short of traditional significance thresholds, we should remember this information relates a conditional effect: the average shift in effect per unit of time. The specific values of both continuous interacted variables – time and margin of victory – will affect significance of the combined effect. Before abandoning the hypothesis, we should chart electoral security's effect at multiple values of both variables. The resulting patterns (see Figure 7) demonstrate one circumstance under which coefficients alone may be misleading. Here, the divergent behavior of very secure and very insecure new members has canceled out the aggregate finding.

[FIGURE 7 HERE]

The shaded bars in Figure 7 indicate the interval between the upper and lower 95% bounds from the simulated sample of relative hazards. I selected a range of interesting values of security, victory by 5%, 25%, 75%, and 99%.[32] As mean-centering sets the average to zero the relative hazard will compare these values to the mean. As Meinke (2005, 109-113) expected, congresspersons enjoying a wider than average

margin of victory are more likely to switch their positions. However, the approximately 50% increase in hazard, which he reported, does not hold over time. The negative trend rapidly pulls the combined effect into statistically insignificant territory, from which it does not re-emerge. The original argument, then, receives qualified support: safer representatives are more likely to switch their votes *only* very early in their careers. Moreover, highly insecure congresspersons exhibit the opposite behavior, with initial relative hazards half the size of those at the mean level of security.

[FIGURE 8 HERE]

Figure 8 provides an alternative depiction, charting the first differences of two hypothetical changes in electoral security. The solid black line gives the median percentage difference in hazard of switching for an increase from a very comfortable win to an uncontested race. The solid gray line marks the corresponding difference were the candidate's margin to decrease from the comfortable win to a victory by plurality.[33] The thin, dashed lines of corresponding color mark the 5th and 95th percentiles from a simulated distribution of 1,000.[34] Running uncontested increases the probability of switching by 20% compared to a comfortable victory. The displayed difference for a move from comfort to tough competition shows a symmetrical decrease of 20%.[35] For both cases, the impact of vote margin "wears off" quickly, dropping out of significance within less than six months.

The establishment of reputations and voting records may produce this dynamic pattern. Super comfortable new representatives may not fear electoral reprisal, but over time they become reticent to change position, viewing continued success as support for their policies (Meinke 2009, personal communication). For an embattled new

representative, establishing a reputation for honesty and holding onto the existing base will dominate strategy, producing the lower initial likelihood of switching positions.

6. Conclusion

The replications above address a wide selection of substantive interests. The link between legislative deliberations, civil war duration and vote choice may seem obscure. It is, however, deep and fundamental: all three describe a political process in which the causal role of explanatory factors is dynamically contingent. As political scientists, we are inherently interested in how these and other complex interactions unfold over time. The NPH Cox model provides an opportunity to develop statistical tests better suited to such social processes, but it requires the commitment of additional post-estimation effort. In each replication study, such an effort added richness to key findings and more thorough evaluation of hypotheses. Appropriate application of these hazard ratio, relative hazards and first differences strategies will provide the necessary follow-through for scholars of civil war, legislative bodies, political position-taking, and the myriad other time-dependent processes fundamental to political science research.

Works Consulted

Ai, Chunrong and Edward C. Norton. 2003. "Interaction Terms in Logit and Probit Models". *Economics Letters*, 80: 123-29.

Balch-Lindsay, Dylan, Andrew J. Enterline, and Kyle A. Joyce. 2008. "Third-Party Intervention and the Civil War Process." *Journal of Peace Research*, 45(3): 345-63.

Box-Steffensmeier, Janet and Christopher J.W. Zorn. 1998. "Duration Models and Proportional Hazards in Political Science, v 1.9". Paper prepared for presentation at the Annual Meeting of the Midwest Political Science Association, April 23-25, 1998, Chicago, Illinois. Available at <http://polmeth.wustl.edu/workingpapers>.

Box-Steffensmeier, Janet and Christopher J.W. Zorn. 2001. "Duration Models and Proportional Hazards in Political Science". *American Journal of Political Science*, 45(4): 972-88.

Box-Steffensmeier, Janet, Dan Reiter and Christopher J.W. Zorn. 2003. "Nonproportional Hazards and Event History Analysis in International Relations". *Journal of Conflict Resolution*, 47(1): 33-53.

Box-Steffensmeier, Janet and Bradford S. Jones. 2004. *Event-History Modeling: A Guide for Social Scientists*. Cambridge: Cambridge University Press.

Brambor, Thomas, William Roberts Clark and Matt Golder. 2006. "Understanding Interaction Effects: Improving Empirical Analyses". *Political Analysis*, 14(1): 63-82.

Chiozza, Giacomo and H.E. Goemans. 2004. "International Conflict and the Tenure of Leaders: Is War Still *Ex Post* Inefficient?" *American Journal of Political Science*, 48(3): 604-619.

Cox, Gary W. 1997. *Making Votes Count: Strategic Coordination in the World's Electoral Systems*. Cambridge: Cambridge University Press.

Downs, Anthony. 1997. *An Economic Theory of Democracy*. Reading, MA: Addison Wesley.

Duverger, Maurice. 1954. *Political Parties: Their organization and activity in the modern state*. Indianapolis: Wiley.

Friedrich, Robert J. 1982. "In Defense of Multiplicative Terms in Multiple Regression Equations." *American Journal of Political Science*, 26(4): 797-833.

Golub, Jonathan. 2007. "Survival Analysis and European Union Decision-Making". *European Union Politics*, 53(4): 733-64.

Golub, Jonathan and Bernard Steunenberg. 2007. "How Time Affects EU Decision-Making". *European Union Politics,* 8(4): 555-66.

Herron, Michael C. 1999. "Postestimation Uncertainty in Limited Dependent Variable Models." *Political Analysis* 8: 83-98.

Inglehart, Ronald and Christian Welzel. 2005. *Modernization, Cultural Change and Democracy: The Human Development Sequence.* Cambridge: Cambridge University Press.

Kalbfleisch, John D. and Ross L. Prentice. 1980. *The Statistical Analysis of Failure Time Data*. New York: John Wiley and Sons.

Kam, Cindy D. and Robert J. Franzese, Jr. 2007. *Modeling and Interpreting Interactive Hypotheses in Regression Analysis*. Ann Arbor: University of Michigan Press.

Kitschelt, Herbert. 1989. *The Logics of Party Formation: Ecological Politics in Belgium and West Germany*. Ithaca: Cornell University Press.

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science*, 44: 347-361.

Light, Paul C. 1999. *The President's Agenda: Domestic Policy Choice from Kennedy to Clinton*, 3rd Edition. Baltimore: Johns Hopkins University Press.

Meinke, Scott R. 2005. "Long-Term Change and Stability in House Voting Decisions: the Case of the Minimum Wage". *Legislative Studies Quarterly*, 30(1): 103-126.

Mooney, Christopher Z. 1997. *Monte Carlo Simulation*. Thousand Oaks: Sage University Press.

Toynbee, Arnold. 1972. *A Study of History*. Oxford: Oxford University Press.

Warwick, Paul V. 1992. "Rising Hazards: An Underlying Dynamic of Parliamentary Government". *American Journal of Political Science*, 36(4): 857-76.

Wittrock, Jill. 2008. "Voting behavior after electoral reform: Cross-national experiments". Ph.D. dissertation, The University of Iowa, 190 pages.

Wright, Quincy. 1942. *A Study of War*. Chicago: University of Chicago Press.

Zhelyazkova, Aysa and Rene Torenvlied. 2009. "The Time-Dependent Effect of Conflict in the Council on Delays in the Transposition of EU Directives." *European Union Politics*, 10(1): 35-62.

Table 1. Replication of NPH Cox Model of EU Decision-Making Speed in Golub (2007)

Qualified Majority Voting (QMV)	3.1218*** (0.4784)
QMV After Single European Act (SEA)	2.1096*** (0.5127)
QMV After Maastricht	0.4134** (0.1657)
Cooperation Procedure	-6.0408*** (0.6135)
Codecision Procedure	-5.0008*** (0.8763)
EU with 9 members	0.4962** (0.1984)
EU with 10 members	0.4569* (0.2429)
EU with 12 members	0.6589** (0.2569)
EU with 15 members	0.5706** (0.2625)
Thatcher (as Prime Minister)	-1.7162*** (0.3794)
Expanded Legislative Agenda	0.1774 (0.1914)
Legislative Backlog	0.0260*** (0.0066)
QMV×ln(t)	-0.4281*** (0.0793)
QMV after SEA×ln(t)	-0.2239*** (0.0846)
Cooperation×ln(t)	0.8904*** (0.0992)
Codecision×ln(t)	0.7251*** (0.1342)
Thatcher×ln(t)	0.2819*** (0.0611)
Legislative Backlog×ln(t)	-0.0040*** (0.0009)

NOTE: 3,001 observations of 1,669 separate pieces of legislation. Data are right-censored on 17 December 1999. For more information on variables see Golub (2007). Standard errors in parentheses.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 2. Replication of Competing Risks Cox Model of Civil War Termination in Balch-Lindsay et al. (2008)

	Government Victory	***Opposition Victory***	***Negotiated Settlement***
Intervention Supporting Government	2.49† (1.83)	1.93 (1.63)	5.67† (3.79)
*Intervention for Government*ln(time)*	-0.63** (0.27)	-0.43** (0.23)	-0.75† (0.50)
Intervention Supporting Opposition	-1.64 (1.12)	1.74*** (0.48)	1.39** (0.56)
Balanced Intervention	1.748† (1.34)	-32.60*** (1.06)	-2.63** (1.21)
Separatist	0.87 (1.05)	0.10 (1.71)	8.16*** (2.93)
*Separatist*ln(time)*	-0.22 (0.18)	-0.18 (0.26)	-1.23*** (0.41)
War Costs	108.67 (204.70)	171.25*** (62.88)	-211.75† (158.26)
*War Costs*ln(time)*	-34.67 (38.05)	-33.56*** (11.80)	25.08 (20.40)
GovReputation	-1.30*** (0.51)	-0.66 (0.58)	0.16 (0.62)
Economic Development	0.96*** (0.15)	0.93*** (0.22)	0.51† (0.37)
*Economic Development*ln(time)*	-0.17*** (0.02)	-0.18*** (0.04)	-0.08† (0.05)
Binary Measure of Democracy	0.33 (0.33)	-0.34 (0.53)	-1.19 (1.22)
Civil Wars	213	213	213
Civil War Failures	109	45	40
Time at risk	249,462	249,462	249,462
Spells	924	924	924
Log-Likelihood	-427.88	-174.45	-132.85
Wald $\chi^2_{(12,12,12)}$	80.54***	3,393.58***	36.16***

NOTE: Data are from Balch-Lindsay, Enterline and Joyce (2008). Figures reported are coefficients from the Cox model. Positive coefficients indicate a decrease in duration (increase in risk); negative numbers an increase in duration (decrease in risk). Robust standard errors are reported in parentheses.

Significance (two-tailed): ** significant at 5%; *** significant at 1%
Significance (one-tailed): † significant at 10%

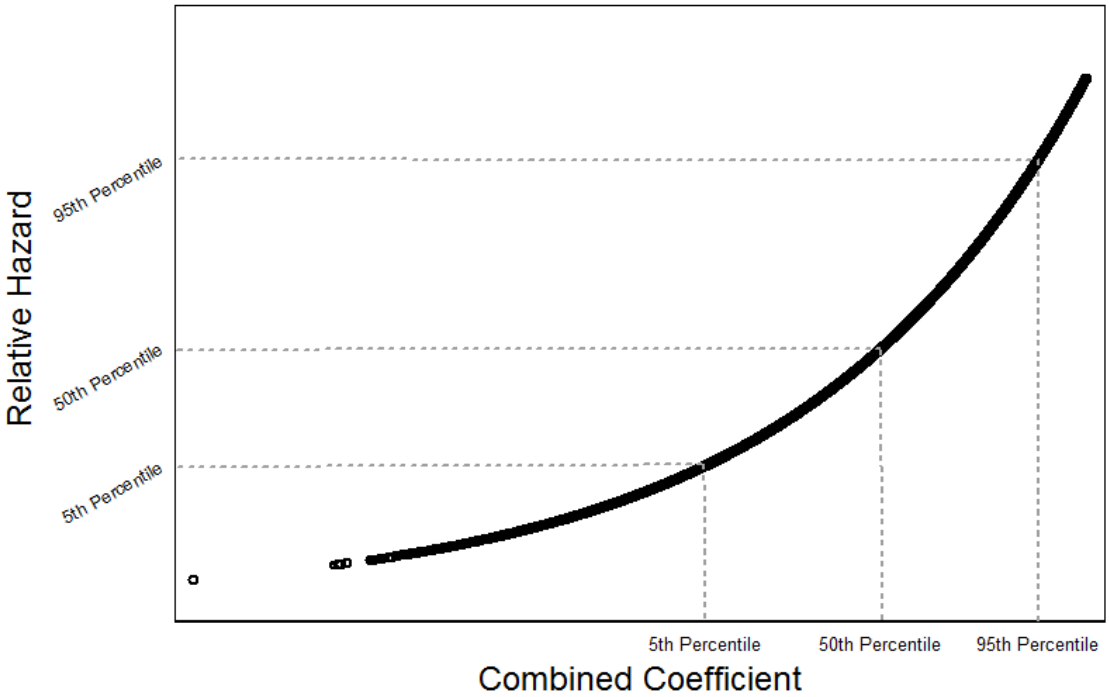
Table 3. Replication of NPH Cox Model of Vote Switching on Issue of Minimum Wage in the U.S. Congress in Meinke (2005)

Party (1=Democrat)	-1.4250*** (0.1283)
Crosspressuring (1=Crosspressured)	0.6978*** (0.1691)
White House Change to Member's Party	0.0695 (0.1632)
What House Change from Member's Party	0.7832*** (0.1747)
Union Membership	-0.0365*** (0.0064)
Size of Wage Change	0.6192*** (0.1883)
Member's Share of 2-Party Vote	1.1965** (0.5165)
WH Change from Member's Party×ln(t)	-0.5088** (0.2171)
Member's Share of 2-Party Vote×ln(t)	-0.9926 (0.6287)

NOTE: N =2,798 congressperson-years with 1,216 subjects and 312 observed failures. Beta values are coefficients for a Cox proportional hazards conditional risk set model with time measured as time from entry into the data set. Baseline hazard rates stratified by order of failures (failures beyond third combined with third strata). Efron method used for tied observations. Robust standard errors in parentheses.
Union membership and vote share are mean-centered variables.

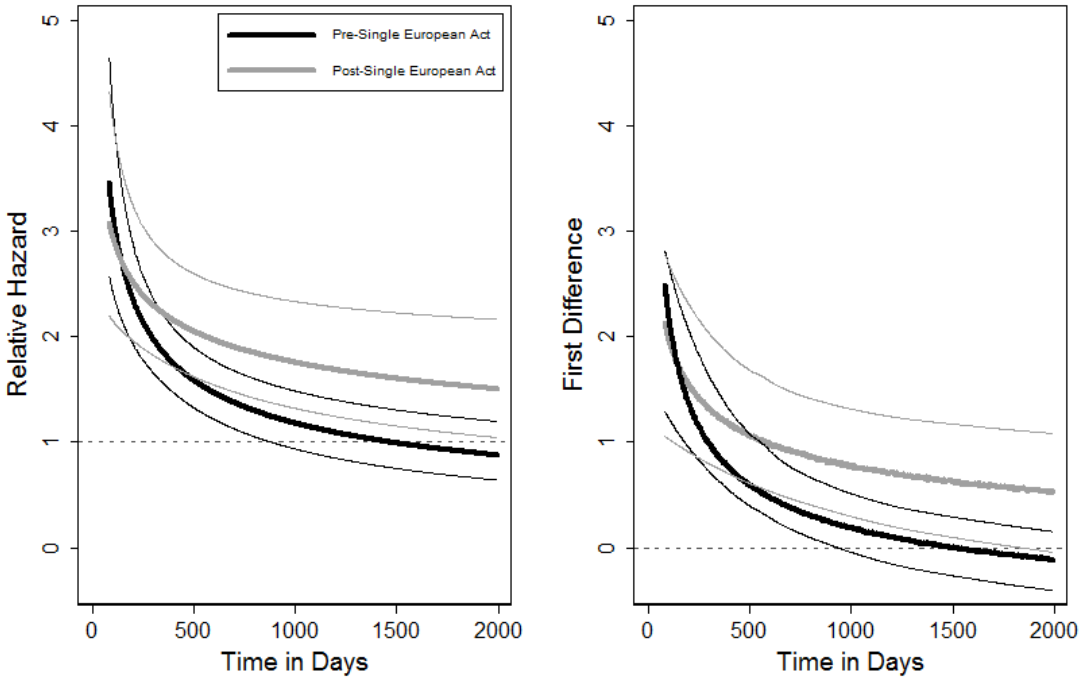
* significant at 10%; ** significant at 5%; *** significant at 1%

Figure 1. Relationship between Distribution of Combined Coefficient and Relative Hazard



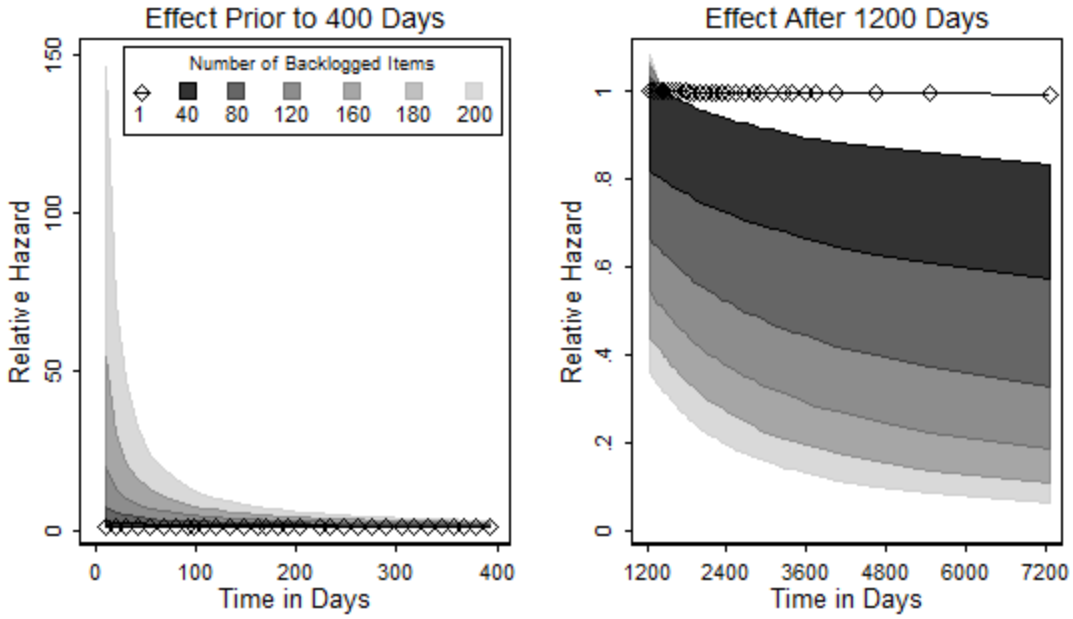
NOTE: Simulated data with draw of 500 pairs of beta1 and beta2 from a normal distribution with mean vector [.5, -.2] and variance matrix [.04, -.001\-.001, .02].

Figure 2. Effect of Qualified Majority Voting



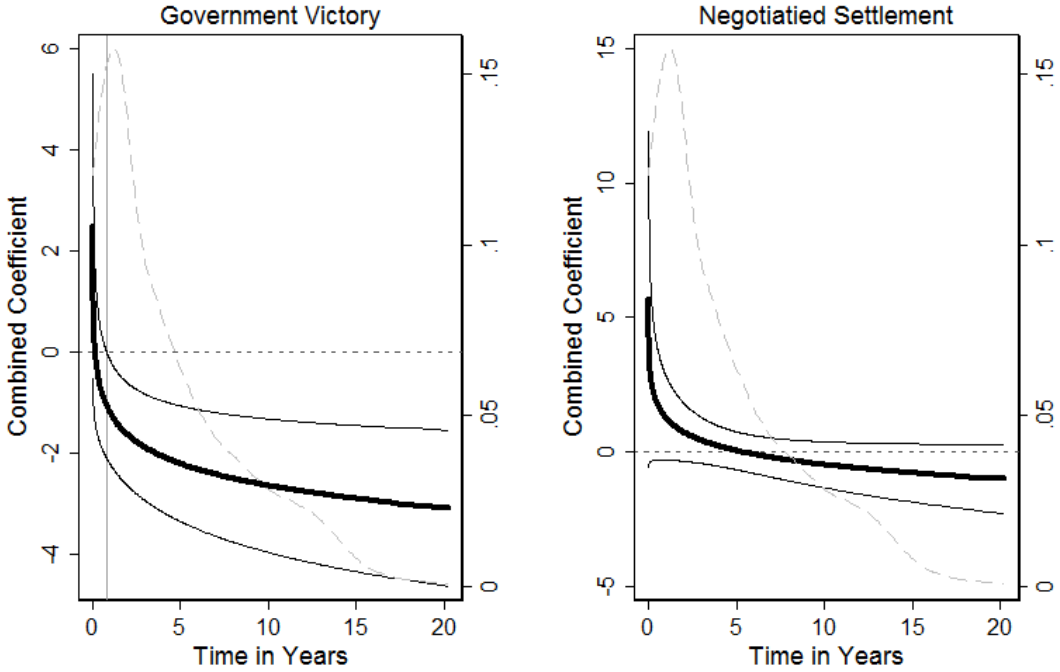
NOTE: Thin lines of corresponding color indicate 95% confidence interval. In order to demonstrate similarity to the relative hazard for binary variables with PHA-violating effects, the first difference is not reported in terms of percentage change. The mean difference from a simulated distribution of 1,000 draws from the parameter matrices of the NPH Cox model in Table 1 is reported, with confidence intervals taken from the same distribution. The confidence bounds were smoothed using a lowess function.

Figure 3. Relative Hazard of Legislative Backlog
Across its range and Over Time



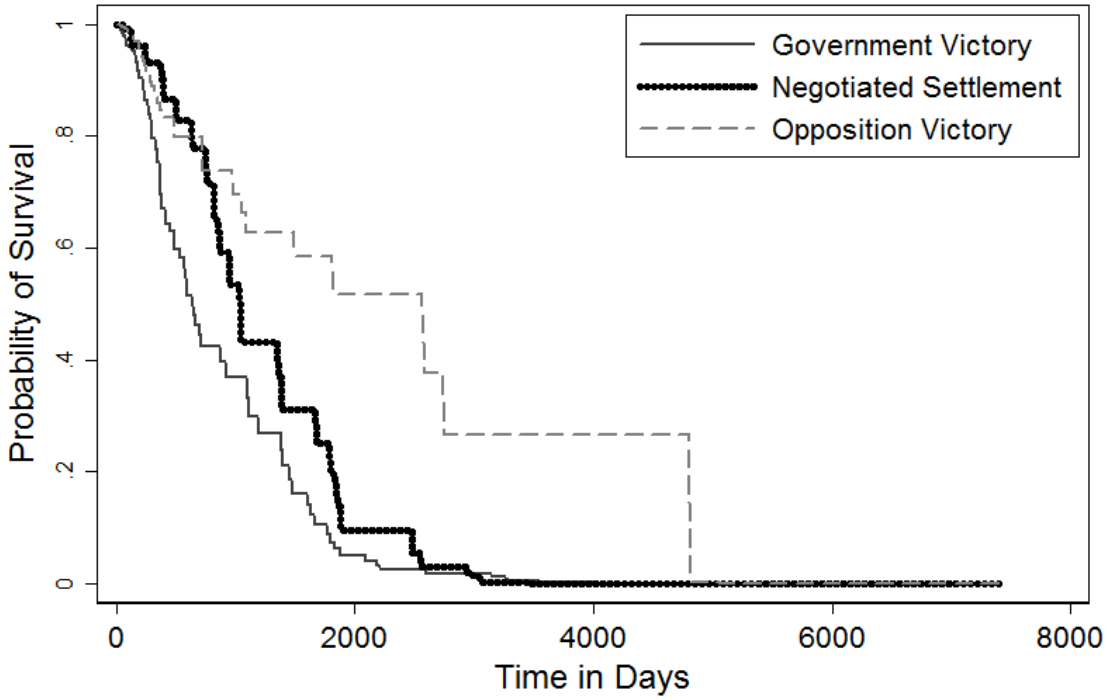
NOTE: Shaded bars represent the 95% confidence intervals based on ten draws of 1000 coefficient estimates for each value of legislative backlog and time. The mean number of backlogged items is 169, minimum 37 and max of 229. Black diamonds mark the effect with one backlogged item, which is that evaluated by Golub and Steunenberg (2007). During the period between day 300 and day 1400 effects at all values are statistically insignificant. The majority of this range is omitted to ease observation of the effect while it is significant.

Figure 4. Combined Coefficient of Pro-Government Intervention



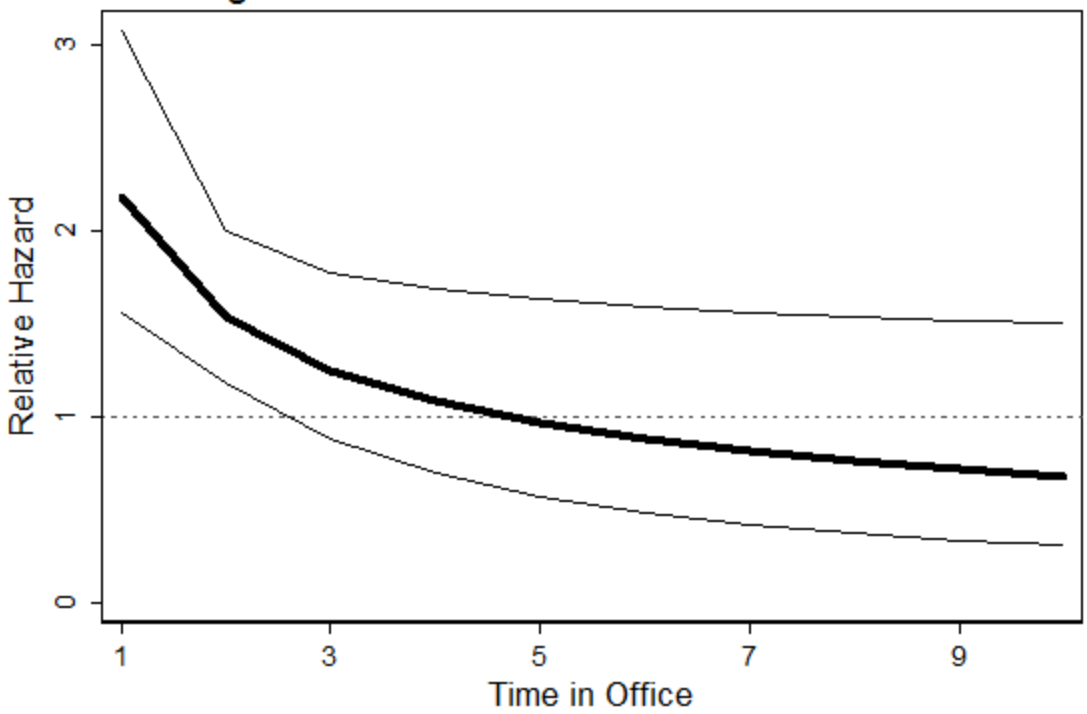
NOTE: Thick line charts the combined coefficient from the appropriate equation of the competing risks NPH Cox Model reported in Table 2. Thin lines mark 90% confidence intervals and dashed curve gives distribution of civil war durations. The thin vertical gray line in the Government Victory graph marks 286 days, the point when the effect becomes significant. 85% of civil wars endure beyond this point in time. Dashed gray line highlights zero. Y-scales differ across panels.

Figure 5. Survival Probability Plots for Civil War Outcomes Given External Assistance to Government



NOTE:Replication of Figure 1 from Balch-Lindsay et al. (2008) using replication files provided by the authors. Each survival curve was derived from a separate NPH Cox model with dependent variable recoded for competing risks.

Figure 6. Relative Hazard of White House Loss

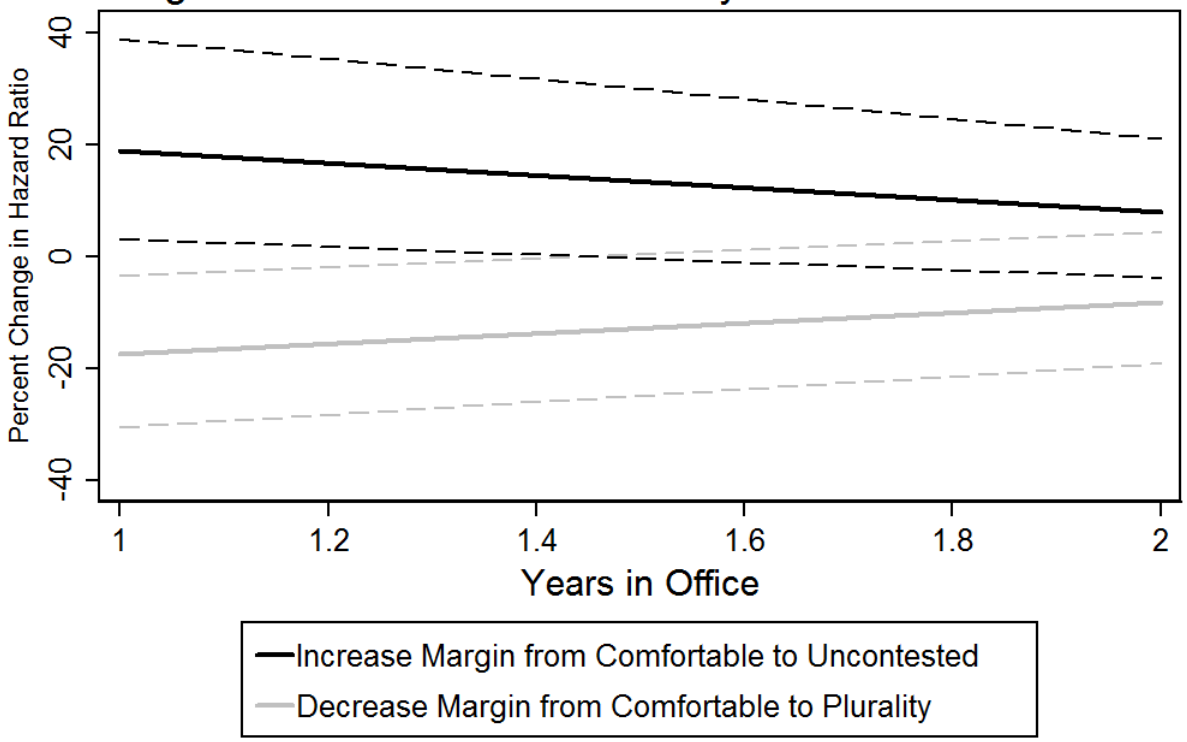


NOTE: Thick line marks the relative hazard. Thin lines indictate 95% confidence intervals. The effect is insignificant when the confidence intervals include 1, which occurs between 2 and 3 years in office. 67% of cases experience a significant positive effect on the likelihood of switching votes when their party loses the White House.

Figure 7. Relative Hazard of Vote Share

NOTE: Shaded bars represent 90% confidence intervals around the median estimate from 1000 draws from the parameter matrix of the NPH Cox model in Table 3. Vote share averages 69.4%, with minimum of 0 and maximum of 100%. Mean centered variable used in estimation and calculation. The effect is significant when its confidence intervals do not include 1.

Fig 8. First-Differences Summary of Vote Share's Effect



NOTE: Thin lines indicate 90% confidence interval around the median calculated change based on a draw of 1000 estimates from the variance-covariance matrix of NPH Cox model in Table 3. Vote share averages 69.4%, with minimum of 0 and maximum of 100%. 'Comfortable' here is one standard deviation about the mean, or a vote share of 85.3% of the vote. The max of 100%, indicates an uncontested race. 'Plurality' is two standard deviations below the mean, a share of 37.3%. Mean-centered variable used in estimation and calculation. The effect is significant when confidence intervals do not include 0.

[1] The simulation procedures are easily programmed. An adaptable Stata do-file is available in Appendix II of supplementary materials on the author's website.

[2] Several very helpful contributions from political methodologists exist to aid in the interpretation of interaction effects in linear regression, limited dependent variable models, some count models and some fully parameterized duration models. These include King, Tomz and Wittenberg's (2000) Clarify software, Boehmke's (2006) Stata utility – `grinter` – for calculating marginal effect of interacted variables, and the Stata code written by Brambor, Clark and Golder (2006).

[3] As will be clarified in the examples below, the vital aspect here is not that the effect of a variable is changing, but that the changing effect of a covariate over time is not *proportional* to the effects of other variables.

[4] Similar patterns of changing loyalty and support for the executive have been located in Western Europe (e.g. Warwick 1992).

[5] By initial effect, I mean that when t=1 and ln(t)=0.

[6] Detailed expositions of the meaning and interpretation of coefficients and standard errors in multiplicative-interaction models can be found in Freiderich (1982), Brambor, Clark and Golder (2006), and Kam and Franzese (2007).

[7] In nonlinear modeling, however, the marginal effect of X_1 will be conditioned by the value of all included **X**, not just that with which it is explicitly interacted (Kam and Franzese 2007, 112; Ai and Norton 2003).

[8] This computer code is freely available at <http://homepages.nyu.edu/~mrg217/interaction.html>. Kam and Franzese's (2007) equivalent code is available at the University of Michigan Press page for their book *Modeling and Interpreting Interactive hypotheses in Regression Analysis*, as well as at the authors' personal web pages, including, <http://www-personal.umich.edu/~franzese/Publications.html>.

[9] To install grinter, open Stata and type the following command:

```
net from http://myweb.uiowa.edu/fboehmke/stata/grinter
```

[10] For more detailed discussion of the recovered baseline hazard, see Box-Steffensmeier and Jones (2004, 65-66).

[11] As there are simpler ways to achieve a substantively interesting means of evaluating the effect of X_i, demonstrating the steps for differentiation is not necessary. Using the product and chain rules, the first derivative of Equation 4.1 with respect to X_i can be shown to be:

$$\partial Y/\partial X = (\beta_1 + \beta_2 \ln(t_i))\, e^{(\beta_1 X_i + \beta_2 \ln(t_i) X_i + \boldsymbol{\beta Z_i})} h_0(t_i) - \sum\left[\beta_1 + \beta_2 \ln(t_i)\, e^{(\beta_1 X_i + \beta_2 \ln(t_i) X_i + \boldsymbol{\beta Z_i})}\right] e^{(\beta_1 X_i + \beta_2 \ln(t_i) X_i + \boldsymbol{\beta Z_i})}$$

[12] In this calculation, and those to come, I denote t as a constant at value t=t. This is vital for the benefits of the relative hazard to kick in: if t is not held constant across the hypothetical cases of i and j, then the baseline hazard will not drop out of the ratio, and a simple quantity will not be located.

[13] This measure is also sometimes called the conditional linear coefficient.

[14] For this illustration, I simulated 1,000 draws of pairs of β_1 and β_2 from a normal distribution with mean values of .5 and -.2, variance of .04 and .02 for each β respectively and covariance -.001.

[15] One could also obtain valid confidence intervals by drawing form the variance-covariance matrix and mean vector of coefficients to simulate a sampling distribution (such techniques will be outlined further below). This method produces equivalent results to the exponentiation of the combined coefficient. Evidence supporting this claim can be found in Table A1 and Figure A1 in the appendix of supplementary materials. The Delta method could also be employed to locate the standard error and confidence intervals around the relative hazard. For the case at hand, however, this method's asymptotically-valid linear approximation is less desirable. Confidence intervals around the relative hazard derived using the Delta method may include nonsensical values. This can be easily verified by using Stata's `predictnl` command to produce confidence intervals via the Delta method.

[16] T will be the number of values of time included in the analysis. If survival times in the original data range from 1-25 years, for example, T could be 25. If the survival times exhibit a much larger or less discrete range, however, say from .1-35.8 years, it may be desirable to create T based on the percentiles of the original distribution. Stata code for either strategy can be found in Appendix II.

[17] For 95% "confidence intervals" this will include the 2.5th, 50th and 97.5th percentiles.

[18] Examination of combined coefficients may also prove useful, but these values must be transformed to lend any substantive appeal.
[19] See Figure 1 and the attending discussion above.
[20] Of course, there is a limit to this logic. A 50% increase in a probability of .0000002, for example, does not significantly alter our expectations regarding the likelihood of the event, and is therefore substantively uninteresting even if it is significant.
[21] For example, the counter may be scaled in years, days, months, minutes, seconds, etc. and the time t=1 will be different depending on which is selected.
[22] Though the estimated effect may flip from positive to negative (or vice versa) at a certain point in time, the effect will of course be statistically indistinguishable from zero for some time before and after this particular point. If desired, the analyst could design a simulation procedure to obtain confidence intervals around the point, or could use the Delta Method. Of course, this information will also be obtained through the process of calculating the full NPH effect with the methods described above.
[23] For more information on the coding and frequencies of these and other variables in the analysis, see Golub (2007).
[24] To facilitate comparison, the differences here were not multiplied by 100. To see the algebraic similarity between these two measures, please review Equations 6 and 8.
[25] The combined coefficient in this case is ($\beta_{backlog} + \ln(t)\beta_{\ln(t)backlog}$).
[26] The effect is insignificant between the 289th and 1,412th days of deliberation.
[27] The deliberation time used here is 2,412 days, which is one standard deviation above the mean deliberation time.
[28] Because intervention is a binary variable, we could calculate the relative hazard and confidence intervals without simulation simply by exponentiation of this combined coefficient. Unfortunately, the insignificant combined impact of pro-government intervention in the early stages of civil war duration produced confidence intervals with extreme values. The chart of the combined coefficient in Figure 4 is more charitable and readable. These extreme values stem from the uncertainty of the finding, not the method of obtaining standard errors. Alternate methods, including the Delta method and simulation procedures, produce analogous results.
[29] I report all findings at the significance level utilized by the original authors. In this case, that means a 90% threshold.
[30]A typographical error in the published manuscript describes the second PHA-violating variable as political party rather than mean-centered vote share. A corrected version of Meinke's (2005) Table 1, as well as replication data, is available online at < http://www.facstaff.bucknell.edu/smeinke/research.htm>.
[31] Two-thirds of observations cluster in the first two years.
[32] These outcomes are not equally likely in the data: 5% and 25% occur twice, 75% occurs 160 times, 99% never. There are, however, a considerable number of uncontested races with vote shares of 100.
[33] The "comfortable win" or "safe" position is one standard deviation above the mean, or a vote share of 85.3%. The "not safe" or "plurality" win is two standard deviations below the mean, or 37.3% of the vote share.
[34] All values below the mean are negative numbers, due to mean-centering. The lower value in this case, then, is added to rather than subtracted from the higher value.
[35] Here, the absolute value of the shift between uncontested and comfortable is roughly have that of the difference between plurality and comfortable (14.7 versus 32.1). The equality of the differences in hazard ratio suggest that behavior is more sensitive to increases in safety than to decreases in safety, as it takes a considerably more dramatic shift downward to produce a "symmetric" change in likelihood of vote-switching.