Duration Models and Proportional Hazards in Political Science

Janet M. Box-Steffensmeier The Ohio State University
Christopher J. W. Zorn Emory University

A key assumption of nearly all widely used duration models is that the hazard ratios (i.e., the conditional relative risks across substrata) are proportional to one another and that this proportionality is maintained over time. Estimation of proportional hazards models when hazards are nonproportional results in coefficient biases and decreased power of significance tests. Techniques for relaxing this assumption allow scholars to test whether the effects of covariates change over time and also permit a more nuanced understanding of the phenomenon being studied. We address the potential problems with incorrectly assuming proportionality, illustrate a number of tests for nonproportionality, and conclude with a discussion of how to accurately and efficiently estimate these models in the presence of nonproportional hazards. We investigate the proportionality assumption for Cox's semiparametric model in the context of the "liberal peace" debate, using data on international conflict in the postwar period.

In recent years political scientists have increasingly adopted a wide range of techniques for modeling duration data. But while the use of duration models by political scientists has increased dramatically, a concomitant examination of the modeling assumptions underlying these methods has not accompanied this growth. An important characteristic of most of these models is the assumption that the relative hazards over different covariate values are proportional. This consideration is important because, as has been widely shown in the statistics literature, estimation of proportional hazards models when hazards are, in fact, nonproportional can result in biased estimates, incorrect standard errors, and faulty inferences about the substantive impact of independent variables (e.g., Kalbfleisch and Prentice 1980; Schemper 1992; Collett 1994; Klein and Moeschberger 1997).

But while the proportional hazards assumption is central to the proper estimation and interpretation of these models, it has received little attention by political scientists modeling duration data. This is unfortunate because, absent explicit examinations of the validity of these assumptions, the reader is often unable to infer whether or not the assumption holds for any particular analysis. This difficulty is exacerbated by the fact that, in many circumstances, both substantive theories and empirical data suggest that the assumption itself is of dubious accuracy. In fact, it is often the case that many substantively interesting hypotheses imply time-dependence or other forms of nonproportionality in the conditional probability of failure. Thus, it is important that, as political scientists begin to use these models more often in applied research, they take care to examine the extent to which this assumption is consistent with their data and to be aware of methods for analyzing duration data in which hazards are not proportional.

Janet M. Box-Steffensmeier is Associate Professor of Political Science, The Ohio State University, Columbus, OH 43210-1373 (jboxstef+@osu.edu). Christopher J. W. Zorn is Assistant Professor of Political Science, Emory University, Atlanta, GA 30322 (czorn@emory.edu).

Thanks to Neal Beck, Brad Jones, Lanny Martin and Dan Reiter for helpful comments. Box-Steffensmeier thanks the National Science Foundation for support via a Mid-Career Methodological Opportunities Grant, number SES-0083418, and Zorn thanks the John M. Olin Foundation for a Faculty Fellowship supporting this research. A previous version of this article was presented at the 1998 Annual Meeting of the Midwest Political Science Association. All analyses were performed in Stata 6.0 and S-Plus 2000; data are available from the authors at http://www.emory.edu/POLS/zorn/Data/index.html.

American Journal of Political Science, Vol. 45, No. 4, October 2001, Pp. 972–988

The plan of our article is as follows. We first discuss the proportional hazards assumption in the abstract, showing what it means for our expectations about the impact of covariates on the conditional hazard. We also discuss forms of nonproportionality, considering examples when we might expect deviations from proportionality to be the norm rather than the exception. We go on to address the impact of nonproportionality in Cox's (1972) semiparametric proportional hazards model.[1] We discuss tests for proportionality, both graphical and statistical, and suggest methods for dealing with nonproportionality should it arise, and illustrate these techniques with widely used data on international conflicts. We show how substantive considerations can prompt concerns about nonproportional covariates and how applied researchers can both test and implement remedies for nonproportionality.

The Issue of Proportionality

Consider a general model of failure time in which the outcome of interest is the duration until the occurrence of some event, which we refer to generically as a "failure." We may write the conditional probability of failure (i.e., the *hazard rate*) at time t as:

$$h(t) = \lim_{\Delta t \to \infty} \frac{\Pr\left(t \leq T < t + \Delta t \middle| T \leq t\right)}{\Delta t} = f\left(X\beta\right) \quad (1)$$

Generally, we are interested in the case where the hazard of failure is a function of a set of k covariates X and a coefficient vector β.[2] For illustrative purposes, consider the widely researched example of cabinet durations, and suppose we have two types of countries in our data, type A and type B. Let $h_A(t)$ and $h_B(t)$ be the hazards of failure at time t for countries of types A and B, respectively. If the hazard at time t for a country of type A is proportional to the hazard for a country of type B, we can express this relationship as:

$$h_A(t) = Ch_B(t) \quad (2)$$

for any positive value of t and where C is a nonnegative constant. Models that can be characterized by Equation (2) are known generally as *proportional hazards* (PH) models. The value of C is the hazard ratio, i.e., the ratio of the hazard of failure at any time for a country of type A relative to a country of type B. If $C < 1$, the hazard of failure at t is larger for a country of type A, relative to a country of type B, such that type A will be expected to experience a cabinet dissolution sooner than type B. Conversely, if $C > 1$, type A's cabinets will, on average, be expected to last longer. In either case, the hazard functions for countries of type A and type B will be roughly parallel over the entire range of failure times (Collett 1994, 44–45; Hosmer and Lemeshow 1999, 205–206). Proportional hazards models thus "assume that the hazard functions of all individuals differ only by a factor of proportionality" (Chung, Schmidt, and Witte 1991, 71). In essence, this means that the effects of covariates are constant over time; the effect of an independent variable is to shift the hazard by a factor of proportionality, and the size of that factor remains the same irrespective of when it occurs.

There are a number of instances where, for substantive reasons, we might expect that the assumption of proportional effects would not hold. In biomedical research, a common reason for nonproportional hazards is that treatment effects decrease over time as subjects develop resistances to therapies. Thus, the hazard of death or morbidity for a treatment group, initially lower than that for the control, increases as the study wears on, causing the two hazard rates to converge. Alternatively, hazards may be diverging, as the impact of a treatment grows more pronounced over time.[3] Finally, in some instances, hazards may actually cross. Collett (1994) gives the example of choosing between traditional drug therapy and surgery in cancer treatments: while the initial risk of the surgery is higher due to complications and other factors, the long-run prognosis of those undergoing surgery is better. Thus initial hazards are higher for surgery

[1] It is important to note that most widely used parametric models for duration data, such as the Weibull, also assume that hazards are proportional. Because of the strong assumptions of parametric models about the shape of the hazard, parametric models are not as widely used outside the social sciences as is the Cox model; as a result, tests and remedies for nonproportionality in the parametric context are largely nonexistent. However, while we focus on the Cox model here (and generally prefer the Cox model due to its less restrictive assumptions), the intuition of our discussion applies to parametric models as well. For interested readers, a brief discussion of nonproportionality in the context of the Weibull model is included in Appendix A.

[2] See Beck (1998), Hosmer and Lemeshow (1999), and Box-Steffensmeier and Jones (2001) for useful introductions to hazard rate models.

[3] It is important to note that converging or diverging hazards are not, by themselves, indicative of nonproportional hazards. In fact, converging hazards are required to meet the assumption of proportionality in models where the hazards are decreasing over time. Similarly, in models where hazards are increasing, we would expect proportional hazards to diverge. What is important is the extent to which that convergence or divergence deviates from proportionality.

patients, but those hazards decline, while those for chemotherapy patients increase, over time. The result is that the estimated hazards for the two treatments "cross" at some point during the process time.

In political science, we might ". . . expect that the effect of one or more predictor variables on the hazard rate increases or decreases over time. There may be a number of different explanations for such change to occur, including learning effects, shifts in life-course position, maturational changes, and so on" (Teachman and Hayward 1993, 359). In such circumstances, covariate effects on the hazard of failure are nonproportional: the influence of an independent variable may be greater or smaller, or even change signs, depending on the amount of time that has elapsed for that observation. Consider an example from recent work on international alliances. Bennett (1997) asserts that, because they are often more vague in purpose and make fewer demands on their members, larger alliances will tend to be more durable (i.e., have lower hazard rates) than smaller ones, and his analysis of alliance data bear out this assertion. Theories of institutionalization, however, also suggest that alliances which survive for long periods of time, regardless of their size, tend to be more self-perpetuating and therefore have lower hazard rates. Thus, the actual effect of alliance size may be large early in the life of the alliance, but decrease over time as the alliance become institutionalized (Zorn 2000). If this is the case, then the effect of alliance size is nonproportional, and estimated hazards for large and small alliances will converge over time.

Estimating proportional hazards models when hazards are in fact nonproportional results in biased coefficient estimates and decreased power of significance tests. In particular, misspecified PH models will overestimate the impact of variables whose associated hazards are diverging, while coefficient estimates for covariates in which the hazards are converging will be biased towards zero (Kalbfleisch and Prentice 1980). Schemper (1992) summarizes the consequences of assuming constant hazard ratios when they are not applicable: "For covariates whose hazard ratios are nonconstant over time, the power of corresponding tests decreases because of suboptimal weights for combining the information provided by the risk sets of times where failures occur (Lagakos and Schoenfeld, 1984). For other covariates with constant hazard ratios, testing power declines as a consequence of an inferior fit of the model" (1992, 455). In addition, the extent of this bias can be consequential. Gray (1996), for example, finds that when two treatments have overlapping or crossing hazards, the power of models based on the proportional hazards assumption can be reduced by as much as 90 percent.

Despite its importance, however, the strong proportionality assumption is rarely tested in political science applications.[4] As a result, in most cases we simply do not know if political science data typically violate this assumption. However, research in other areas has concluded that the assumption is "unrealistic in most applications" (Vermunt 1997, 101) and ". . . that violations of the proportionality assumption are the rule, rather than the exception" (Singer and Willett 1993, 186). We advocate that analysts routinely assess this assumption, using the relatively simple tests described below, and implement measures to model nonproportionality if such effects are suspected or uncovered. In the following sections, we outline such tests and techniques for Cox's (1972) proportional hazards model.

Nonproportionality and Cox's Proportional Hazards Model

The proportional hazards model developed by Cox (1972) is a popular and flexible model that does not assume a specific probability distribution for the time until an event occurs.[5] The absence of a need to parameterize time dependency is a significant advantage in most political science applications, since our theories usually do not allow us to specify *a priori* what distribution should be used, and in many cases the parameterization chosen can have a large impact on the substantive conclusions drawn (Larsen and Vaupel 1993). The hazard rate for the Cox proportional hazards model is:

$$h\left(t \middle| X_i\right) = h_0(t)e^{X_i\beta} \tag{3}$$

where $h_0(t)$ is the (unspecified) baseline hazard function and X_i are covariates for individual *i*. Such models are typically estimated via a quasi- or partial-likelihood procedure, in which the term for the baseline hazard is treated as a nuisance parameter and integrated out of the likelihood (Cox 1972; Hosmer and Lemeshow 1999). The Cox model assumes that the hazard functions of any two individuals with different values on one or more covariates

[4] For recent exceptions, see Diermeier and Stevenson (1999) and Martin (2000).

[5] The Cox model is the continuous time analogue to a discrete-time model with a complementary log-log link (Holford 1980; Laird and Oliver 1981). Issues relating to nonproportionality are also of concern in such models; however, because discrete-time models require explicit formulation of duration dependence, the means for addressing such issues are somewhat different than for the continuous-time case; we discuss this further below.

differ only by a factor of proportionality. The baseline hazard rate varies with time but not across individuals, so that the ratio of the hazards for individuals i and j are independent of t and are constant for all t:

$$\frac{h_i(t)}{h_j(t)} = e^{\beta(X_j - X_i)} \tag{4}$$

As noted above, estimation of Cox's model when hazards do not satisfy the proportionality assumption can result in biased and inefficient estimates of *all* parameters, not simply those for the covariate(s) in question. As a result of this possibility, it is widely recognized outside political science that the use of procedures to assess the validity of this assumption are critical.

Tests for nonproportionality in the Cox model fall into three general classes (see Ng'andu 1997 for a recent survey). All can be thought of as variations on a more generalized Cox model that allows hazard ratios to vary over time:

$$h(t) = h_0(t)e^{\left[X_i\beta + (X_i g(t))\gamma\right]} \tag{5}$$

In the specification given in Equation (5), the effects of individual covariates are allowed to vary according to some function $g(\cdot)$ of time. The intuition behind many tests for nonproportionality is to test for $\gamma = 0$. The three general classes of tests that fall into this framework are:

- Tests based on changes in parameter values for coefficients estimated on a subsample of the data defined by t,
- Tests based on generalized regression residuals, and
- Explicit tests of coefficients for interactions of covariates and time.

The last of these approaches also provides a way of explicitly modeling time-dependent covariate effects in the Cox model.[6] We address each of these approaches in turn.

Nonproportionality Tests Based on Piecewise Regressions

The nature of the time-dependence specified in Equation (5) implies that covariates will have differential impacts on the hazard rate depending on the time in the event history at which they occur. The simplest possibility is to treat $g(\cdot)$ as a step function at some point in the process τ, taking on a value of 0 for all points in time prior to that and 1 after:

$$\begin{aligned} g(t) &= 0 \forall t \leq \tau, \\ &= 1 \forall t > \tau. \end{aligned}$$

Under this view, a natural test for nonproportionality is to estimate separate Cox regressions for different values of $g(t)$, that is, for observations whose survival times fall above and below some predetermined value, and determine if the estimated covariate effects are consistent across the two models.

The piecewise regression approach provides a simple means of making an initial assessment of the issue of proportionality (Schemper 1992; Collett 1994). While better tests are now available for the Cox model (i.e., tests based on the generalized residuals, which we discuss in the following section), for researchers choosing to employ a Weibull or other parametric model that makes the proportional hazards assumption, this test is the best one can do. The specification above represents this test in its simplest form; the data may be divided into as few or many separate time periods as are reasonable and/or are suggested by one's theory. In some cases, aspects of the data will suggest likely values for the break points; otherwise, medians or quartiles may be used. By examining the sensitivity of parameter estimates to estimation over different subsets of the data, one can implicitly test the hypothesis that each covariate's impact remains relatively stable over the period under study.[7]

Residual-Based Tests for Proportional Hazards

The second set of approaches for detecting violations of the proportional hazards assumption are *residual-based* approaches; these include both graphical and statistical tests for nonproportionality.[8] In standard least-squares

[6]Importantly, we advocate that the latter approach *not* be used as a test, but rather as a means of correcting for nonproportionality, for reasons discussed below.

[7]An alternative means of dealing with nonproportionality is to stratify the data by the covariate of interest. Under stratification, the impact of the remaining independent variables on the conditional hazards is assumed to be constant across strata, but separate baseline hazards are estimated for the j different groups defined by the covariate in question:

$$h(t) = h_{0j}(t)e^{X_i\beta}$$

One benefit of stratification is that it allows straightforward comparisons of model fit and parameter sensitivity to the stratification technique. A major drawback of the stratification approach, however, is that stratification on a variable of interest prevents estimation of the impact of that variable on the hazard rate. Because the Cox model factors out the baseline hazard, the separate baselines for different covariate values are not reported. Thus, stratified estimates tell us nothing about the effect of the stratifying variable on the hazard of failure.

[8]Another graphical method for assessing nonproportionality is the use of log-log plots. These plots have become one of the most widely used methods, largely because they are relatively simple to generate and interpret (e.g., using Stata's -stphplot- command) (Chen and Wang 1991; Deshpande and Sengupta 1995; Grambsch

regression, a residual is simply the difference between the observed value of the outcome variable and its predicted value. Residuals are not as obvious in the context of duration models, since the value of the outcome variable may be censored and the fitted model may not provide an estimate of the systematic component of the model due to the use of the partial likelihood (Hosmer and Lemeshow 1999, 197–198).

The key intuition for comprehending residuals in the context of duration models is to understand the Cox model as a special case of a more general "counting process." This conceptualization was first suggested by Andersen and Gill (1982); while analytically difficult, it provides a very general, unified way of conceptualizing models for duration data, and has been widely adopted in the biomedical literature (see, generally, Andersen et al. 1993; Fleming and Harrington 1991). Under this view, "each subject in the data is treated as one observation in a (very slow) Poisson process. A censored subject is thought of not as *incomplete data,* but as one whose event count is still zero" (*Guide to Statistics* 1999, 277). Under this approach, we can define the residual, $\hat{M}_i(t)$, as simply the difference between the "observed" event indicator (i.e., the censoring indicator $\delta_i(t)$) and the "expected" number of events (i.e., the integrated hazard estimate $\hat{\Lambda}_i(t) = \int_0^t h_i(u)du$):

$$\hat{M}_i(t) = \delta_i(t) - \hat{\Lambda}_i(t) \tag{6}$$

More specifically, Fleming and Harrington (1991, 163–197) derive residuals for the Cox proportional hazards model by considering that model as a special case of this more general multiplicative intensity counting model. They show that the estimated martingale residual $\hat{M}_i(t)$ for the nontime-varying Cox model is:

$$\hat{M}_i(t) = \delta_i(t) - \int_0^t \exp\left(X_i\hat{\beta}\right) d\hat{\Lambda}_0(t) \tag{7}$$

where $\delta_i(t)$ is a censoring indicator and $\hat{\Lambda}_0(t)$ is the estimate of the integrated baseline hazard at t. These residuals have several properties reminiscent of ordinary least-squares residuals: for example, $\Sigma M_i = 0$, and $\mathrm{Cov}(M_i, M_j) = 0$ asymptotically.[9] Additionally, Therneau, Grambsch, and Fleming (1990, 151) note that covariate-specific score residuals can be derived by using the martingale residuals and considering the derivative of the Cox partial likelihood with respect to each coefficient β_k:

$$L_{ki}(t) = \int_0^\infty \left[X_{ki}(t) - \overline{X}_k(t)\right] d\hat{M}_i(t) \tag{8}$$

where $\overline{X}_k(t)$ is the weighted mean of covariate $\mathbf{X}_k$ over the risk set at time t, with weights corresponding to $e^{X_k(t)\beta}$. In contrast to martingale residuals, score residuals are covariate specific, and sum to zero across observations for each covariate. Score residuals are closely related to the partial residuals introduced by Schoenfeld (1982), which have been widely used to test the proportional hazards assumption. Specifically, the Schoenfeld residuals for each covariate k are simply the cross-observation sums of the efficient score residuals:

$$s_{kt}(t) = \sum_{i=1}^{N} L_{ki}(t) \tag{9}$$

This summation yields a single value for each covariate at each time point, which can then be used to diagnose violations of the critical proportional hazards assumption.

Fleming and Harrington (1991) illustrate how martingale, deviance, score, and Schoenfeld residuals may be used for assessing model adequacy, including testing the assumption of proportional hazards. More recent work by various authors (e.g., Grambsch and Therneau 1994; Grambsch, Therneau, and Fleming 1995) has extended these techniques considerably. Schoenfeld residuals have emerged as particularly important in testing the proportional hazards assumption, in two ways. First, if the hazards are proportional, the Schoenfeld residuals should be a random walk over the range of survival times; that is, there should be no relationship between an observation's residual for that covariate and the length of its survival time. Conversely, if proportional hazards does not hold, the fitted model will underestimate the hazard during those periods where the hazards are diverging

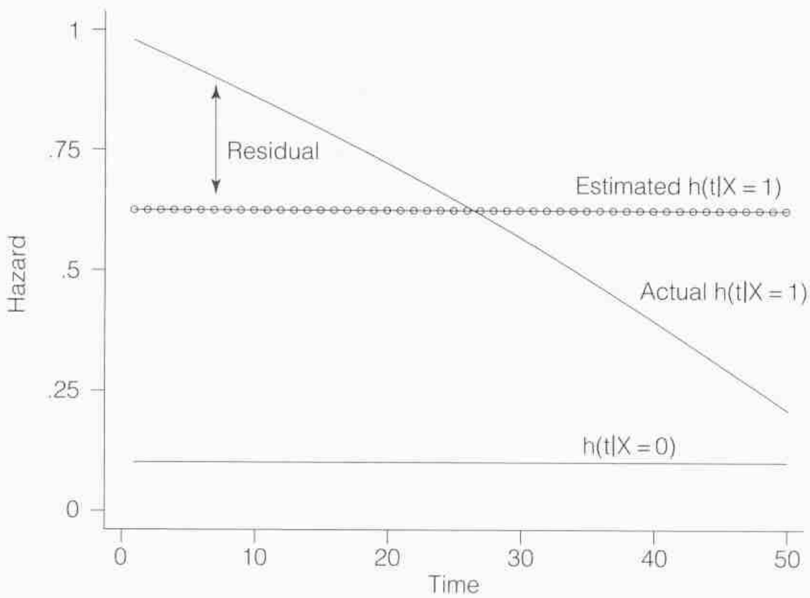
et. al. 1995). However, this approach has come under substantial criticism for its failure to consistently and correctly diagnose instances of nonproportionality, particularly in the presence of additional covariates, and for its inability to assess nonproportional effects in continuous covariates (e.g., Chastang 1983; Schemper 1992). Accordingly, we do not recommend this method for detecting nonproportionality and do not consider the approach further here.

[9]One potential drawback of the martingale residuals defined in this fashion is that they are badly skewed. In the context of the Cox model, martingale residuals are lower unbounded but bounded from above by one. Therneau, Grambsch, and Fleming (1990) suggest a normalizing transformation of the residuals, similar to the deviance residuals common to generalized linear models, which leads to the residuals being symmetrically distributed around zero when the correct model is estimated (e.g., McCullagh and Nelder 1989). They suggest using *deviance residuals,* defined as:

$$d_i = sign\left[-2\left(\hat{M}_i + \delta_i \ln\left(\delta_i - \hat{M}_i\right)\right)\right]$$

This formulation inflates the martingale residuals close to one, while reducing the magnitude of very large negative values.

FIGURE 1 Schoenfeld Residuals and Proportional Hazards



Note: Figure plots actual (non-proportional, smooth lines) and estimated (proportional, circled line) hazards for a binary covariate with nonproportional effects. Residuals may be interpreted as differences between observed and expected numbers of events; see text for details.

and overestimate it when they are converging. So if, for example, the hazards for a particular covariate are converging over time, the model will underestimate the impact of that variable for small *t*, and overestimate it for large *t*, a fact that will be reflected in the residuals for that covariate. The intuition of this is illustrated in Figure 1, which plots the actual (nonproportional) and estimated (proportional) hazards for a binary covariate which has a nonproportional effect on the hazard rate. Figure 1 illustrates how plots of the values of the residuals against some function of time can serve as a simple graphical test for proportionality in covariate effects.

Second, Schoenfeld residuals form the basis for statistical tests of the nonproportionality assumption. The intuition discussed above has led to Therneau, Grambsch, and Fleming's (1990) residual-based test, which uses the maximum of the absolute value of the summed (over time) Schoenfeld residuals as a global test for nonproportionality in the model. Relatedly, one can calculate the correlation ρ between the Schoenfeld residuals for a particular covariate and the rank of the survival time (Harrell 1986). Grambsch and Therneau (1994) modify this test by using the scaled residuals and also detail a global test for nonproportionality based on the aggregated (across covariates) covariance between the unscaled Schoenfeld residuals and survival time. In every case, residual-based evaluations of proportional hazards have been facilitated by the development of software which makes the generation and analysis of these residuals routine (e.g., in Stata 7.0 and S-Plus 6.0).[10]

Estimating Cox Models In the Presence of Nonproportional Hazards

In the event that covariate effects are nonproportional, there are two well-known and accepted estimation approaches. First, as discussed above, separate Cox models may be used for two or more distinct time intervals. While this approach has the advantage of being both conceptually and computationally simple, it also suffers from several drawbacks. It forces researchers to divide up their time scale, in many cases in an arbitrary manner, and the decision about where to divide the time axis can have a significant influence on the model's results. Moreover, to

[10]Log files and appropriate commands for estimating residual-based tests for nonproportionality using Stata and S-Plus can be found in an appendix available at the *AJPS* website or at http://www.emory.edu/POLS/zorn/Data/B-S&Z2001.html.

the extent that the model contains covariates that are proportional in their effects, such an approach suffers from inefficiency, since separate analyses force the researcher to estimate two or more parameters where only one would be adequate. Thus, while estimating piecewise models is one test (and, in the case of parametric models, the *only* test) for assessing proportional hazards, it is not the optimal method for correcting such a problem.

A second estimation approach, and one which is widely recommended, involves estimating a standard Cox model with the addition of an interaction effect between the offending covariate(s) and some function (often the natural logarithm) of time (e.g., Kalbfleisch and Prentice 1980; Collett 1994). Thus, to test if covariate X has a nonproportional effect, one would include in the model an additional variable $X_i \times ln(Time)$.[11] This amounts to an explicit operationalization of Equation (5) and is a very general way of addressing nonproportionality; in particular, in the case of a binary covariate, this test encompasses all possible alternatives to proportional hazards (*Guide to Statistics* 1999, 309). Significance tests on the interaction term can be conducted in the standard way and constitute a direct test for the proportionality of the covariate's effect; moreover, in the presence of nonproportionality, including such an interaction results in a better-specified model and greater accuracy in assessing covariate effects.

In summary, theoretically informed use of time-by-covariate interactions provides the best means of estimating models in which the covariate effects are not proportional. As a general rule, the recommended residual-based tests should always be conducted, as it is widely accepted that they are the best approach for detecting nonproportionality. If such tests indicate nonproportionality is present, corrections via interactions of time and the offending covariates can then be implemented for those variables that are shown to be substantially nonproportional in their effects. While some have advocated the use of *ln(Time)* interactions for testing nonproportional hazards as well as correcting for it, this is not recommended: correlation among the covariates, including that which is induced by the addition of all *ln(Time)* interactions, can affect conclusions about the presence or absence of proportional hazards (e.g., Grambsch and Therneau 1994).

[11]Other forms of the interaction have been suggested, such as $X_i \times Time$ or $X_i \times (Time)^2$, each of which reflects the possible diversity in the shape of the nonproportionality (and constitutes a different test). However, most applied treatments favor *ln(Time)* interactions. Moreover, simulations by Quantin et al. (1996) and Ng'andu (1997) show that *ln(Time)* "has power nearly as high as or higher than all other commonly used tests to detect reasonable alternatives to proportional hazards" (Hosmer and Lemeshow 1999, 207).

Proportional Hazards and the Study of International Disputes

We use data on international disputes to demonstrate the importance of the proportional hazards assumption and to illustrate procedures and techniques for detecting and addressing violations of proportionality. For comparability with previous work, we analyze widely used data on 827 "politically relevant" dyads during the period from 1950 to 1985 (e.g., Oneal and Russett 1997; Beck, Katz, and Tucker 1998; Reed 2000).[12] Each dyad is observed once for each year it is in the data, for a total of 20,990 observations (an average of 25.4 years per dyad). The variable of interest is the duration until the onset of a militarized interstate dispute between the two nations who make up the dyad. Following previous work, we model the hazard of a dispute as a function of six factors: the level of *democracy* in the dyad, *economic growth*, the presence of an *alliance* between the two nations, whether the two nations are geographically *contiguous*, the ratio of military *capabilities* of the two nations, and the level of intradyadic *trade* (measured as a proportion of GDP).[13] In addition, in light of Beck, Katz, and Tucker's (1998) analyses, we estimate models both with and without a counter for the number of previous conflicts during the 1950–1985 period.[14] Previous research suggests that all of these covariates save those for contiguity and previous disputes will exhibit a negative impact on the hazard of an interstate dispute.

As a theoretical matter, we have reasons to expect nonproportional effects in the international disputes data.

[12]These data are at the core of the "liberal peace" research agenda, which has become one of the most important current research programs in political science and is at the heart of the debates over the relationship between trade and conflict (Russett 1990, 1993; Maoz and Russett 1992, 1993; Ray 1997; Gartzke 1998; Werner 2000b). Our work with these important data meshes with a current concern in the field of international relations over the changing influence of explanatory variables on conflict over time (e.g., Box-Steffensmeier, Reiter, and Zorn 2000; Mansfield and Pollins 2000). Indeed, as a substantive matter, our work here helps to make sense of the myriad disparate findings on the relationship between trade and conflict (e.g., Barbieri 1996; Oneal et al. 1996; Oneal and Russett 1997, 1999; Beck, Katz, and Tucker 1998; Beck 1999).

[13]All variables are operationalized as in Beck, Katz, and Tucker (1998); see Appendix B for details and summary statistics.

[14]Our *Previous Disputes* variable counts the number of disputatious events experienced by the dyad; consistent with Beck, Katz, and Tucker's analysis (1998, Table 2) we also omit dyad-years of continuing conflicts. This approach is a simple way of accounting for multiple conflicts within the same dyad; for a fuller treatment of the issue of repeated events, see Box-Steffensmeier and Zorn (2001).

For example, the pacifying effects of an alliance, strong in the initial years after a conflict, may wane as the peace wears on. Alliances that occur shortly after a dispute reflect a particularly high degree of commitment to peace between those nations; those which persist, or are formed later, may lack the same level of commitment. A similar dynamic, albeit in the opposite direction, may operate vis-a-vis the effects of previous disputes. Social-psychological theories of learning in international relations suggest that the influence of previous events on behavior may wane over time (e.g., Reiter 1996). Accordingly, we might expect that the increased likelihood of conflict associated with previous disputes may decline over time, as nations "forget their differences" and settle into a lasting peace. In both instances, theory suggests that the effect of these influences will be nonproportional over time.

We begin by examining piecewise Cox models[15] in which we divide the time until disputes into those shorter than fourteen years (the median duration in the data) and those equal to or greater than that value; these results, along with the pooled estimates for all observations, are presented in Table 1. Columns one and four essentially replicate the results of Oneal and Russett (1997) and Beck, Katz, and Tucker (1998), respectively, using all observations: democratic dyads exhibit significantly lower hazards of conflict, as do those with high levels of growth and large disparities in military capabilities. The reverse is true for contiguous states, which are more likely to engage in conflict, while the estimated effects for trade, though negative, fail to attain anything approaching statistical significance. And as in other analyses, the pacifying effect of alliances decreases dramatically once previous events are included in the model (Beck, Katz, and Tucker 1998, Table 3).

More interesting, however, are the findings for the piecewise models. While some estimates (e.g., those for democracy and contiguity) remain relatively consistent across the two piecewise models, others exhibit significant variation. The effect of economic growth, for example, is significantly larger for dyads that have experienced extended periods of peace than for those with shorter durations. Similarly, the effects of differences in capabilities also vary across the two piecewise models, though this difference effectively disappears when previous disputes are included. The results for previous disputes themselves also vary across the two models: as one would expect, the influence of previous conflicts on the instant probability of a dispute declines the longer two nations have been at peace. Most striking are the changes in the effects for international alliances: while both point estimates are negative, the effect of alliances in dyads with longer durations is both tiny and imprecisely estimated. Once previous disputes are controlled for, alliances actually exhibit a small, positive influence on the probability of conflict in dyads that have been at peace for more than the median duration. For both models, likelihood-ratio tests allow us to confidently reject the hypothesis that the coefficients are constant across the two estimates ($\chi^2(6) = 18.82$ and 152.44, $p = .004$ and $<.001$ for the models omitting and including previous disputes, respectively).

This initial look at the issue of nonproportionality, then, suggests that several covariates may not have a proportional influence on the probability of an international dispute. In particular, the effects of economic growth, alliances, and previous disputes all appear to vary depending on the duration of the peace. These findings are reinforced by an examination of the residual-based methods of Grambsch and Therneau (1994). We begin by plotting the rescaled Schoenfeld residuals against survival times; as discussed above, a trend in this plot indicates that the Cox model is systematically over- or underpredicting the actual hazards at particular time points, and provides strong evidence of nonproportionality. We supplement this graphical approach with statistical tests for nonproportionality based on these residuals; we use the Harrell (1986) correlation test for individual variables, as well as calculating Grambsch and Therneau's (1994) global test for nonproportionality.

The Grambsch and Therneau plots for the model that omits previous disputes are presented in Figure 2. The vertical axis indicates the values of the residuals, while the horizontal axis plots *ln(Time)*. Plots also include a reference line at zero (the mean of the residual values) as well as a lowess line (span = 0.8) through the residuals to facilitate observation of trends in the residuals. If the effects of a covariate are proportional, the two lines ought to be very close to one another, as the average value of the residuals at any point in time should be zero. This pattern is generally true for the trade variable, and only slightly less so for the democracy measure. The positive slope of the lines for the alliance and capability ratio variables, by contrast, is consistent with our earlier results: both suggest that a model which assumes a proportional effect for those covariates will tend to underpredict the hazard of a

[15]While we do not recommend piecewise estimation of Cox models for reasons previously discussed, we present these results to illustrate how one can test for nonproportionality in a parametric (e.g., Weibull) model. To the extent that the residual based tests shown below are not available for the parametric models, piecewise regressions provide the most direct way of assessing nonproportionality in these contexts. Here, the piecewise models provide additional complementary evidence of nonproportionality for the Cox model as well.

TABLE 1 Cox and Piecewise Cox Models of International Disputes, 1950–1985

	No Previous Disputes			With Previous Disputes		
Variable	All Observations	T < 14	T ≥ 14	All Observations	T < 14	T ≥ 14
Democracy	–0.439**	–0.485**	–0.333*	–0.333**	–0.247*	–0.245
	(0.123)	(0.142)	(0.176)	(0.108)	(0.132)	(0.185)
Economic Growth	–3.227**	–1.427	–6.187**	–2.702*	–0.778	–5.301**
	(1.318)	(1.718)	(1.763)	(1.331)	(1.945)	(1.756)
Alliance	–0.414**	–0.642**	–0.035	0.110	0.036	0.465*
	(0.170)	(0.190)	(0.243)	(0.114)	(0.141)	(0.216)
Contiguity	1.213**	1.061**	1.486**	0.449**	0.472*	0.583**
	(0.178)	(0.189)	(0.267)	(0.124)	(0.212)	(0.228)
Capability Ratio	–0.214**	–0.262**	–0.137	–0.162**	–0.099**	–0.116
	(0.082)	(0.095)	(0.097)	(0.059)	(0.041)	(0.107)
Trade	–13.162	–13.304	–17.600	11.487	–3.677	5.483
	(13.827)	(30.906)	(13.353)	(6.670)	(19.541)	(6.079)
Previous Disputes	—	—	—	1.062**	1.667**	0.826*
				(0.078)	(0.099)	(0.066)
*ln*L	–2501.88	–1698.52	–793.95	–2094.03	–1367.16	–650.65
N	20,448	10,366	10,082	20,448	10,366	10,082

Note: Cell entries are coefficient estimates; robust standard errors are in parentheses. One asterisk indicates $p < .05$, two indicate $p < .01$ (one-tailed). See text for details.

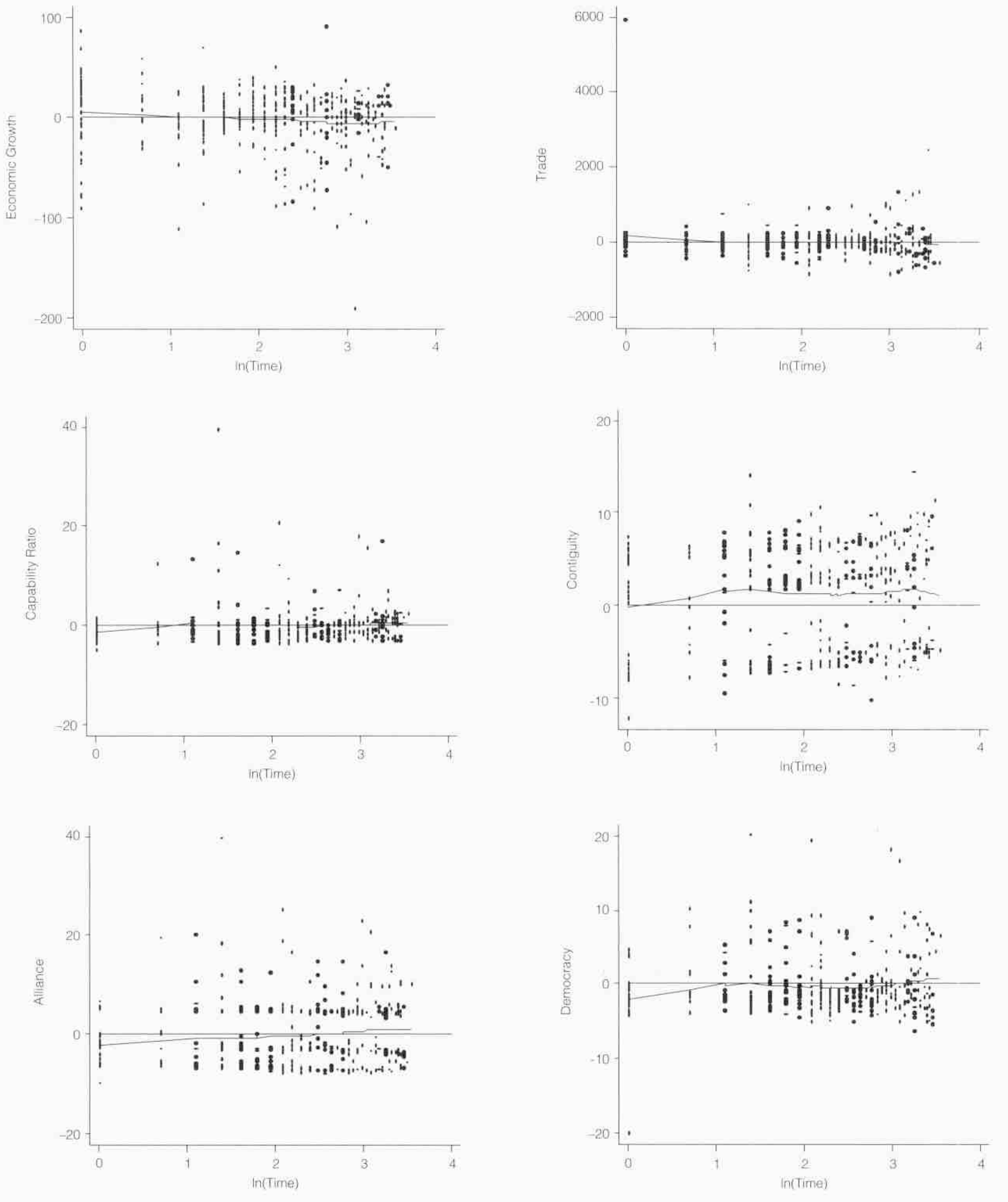
dispute early (when those variables' negative effects are greatest) and overpredict that hazard later in the duration of the peace. The reverse is true for the economic growth variable; the observed negative slope indicates a tendency to overestimate its effects early in the duration and underestimate them in later periods. In all three instances, these patterns are consistent with nonproportional effects for these variables.

Evidence of nonproportionality is even clearer in the residual plots from the model that includes previous disputes as a covariate. Figure 3 again plots scaled Schoenfeld residuals against *ln*(*Time*) for four representative independent variables. The effect of economic growth is once again only slightly nonproportional, while that for levels of trade is again very nearly proportional. By contrast, the nonproportional effect of alliances is even greater once previous disputes are taken into account, as evidenced by the clear positive slope to the scaled residual plot. And the effect of previous disputes themselves are also strongly nonproportional: *all* scaled residuals through T = 5 take on positive values, and the overall slope of the line is clearly negative, indicating that the effect of previous disputes "wears off" after many years of peace.

The formal statistical tests based on Schoenfeld residuals are straightforward and reinforce the findings of the graphical analyses. Table 2 presents the results of both covariate-specific and global Grambsch and Therneau (1994) tests for nonproportionality described above, for models with and without *Previous Disputes.* The columns designated ρ report the estimated correlation between the scaled residuals and *ln*(*Time*), while the χ^2 and p-values indicate the confidence with which we can reject the null hypothesis that the hazard ratios for different values of that covariate are constant over time. For the model without previous events, the clearest evidence of nonproportionality is found in the measures for alliances and capability differences, with smaller effects for democracy, growth, and trade; all these results are consistent with earlier findings for this model. When previous disputes are accounted for, both overall levels of nonproportionality and many variable-specific values also increase; most notably, that for alliances and economic growth are larger in the latter model. The largest nonproportional effects are again observed for the previous disputes variable itself, and overall levels of nonproportionality are significantly higher for the latter model as well; this latter finding is also consistent with the likelihood-ratio tests from the piecewise models.

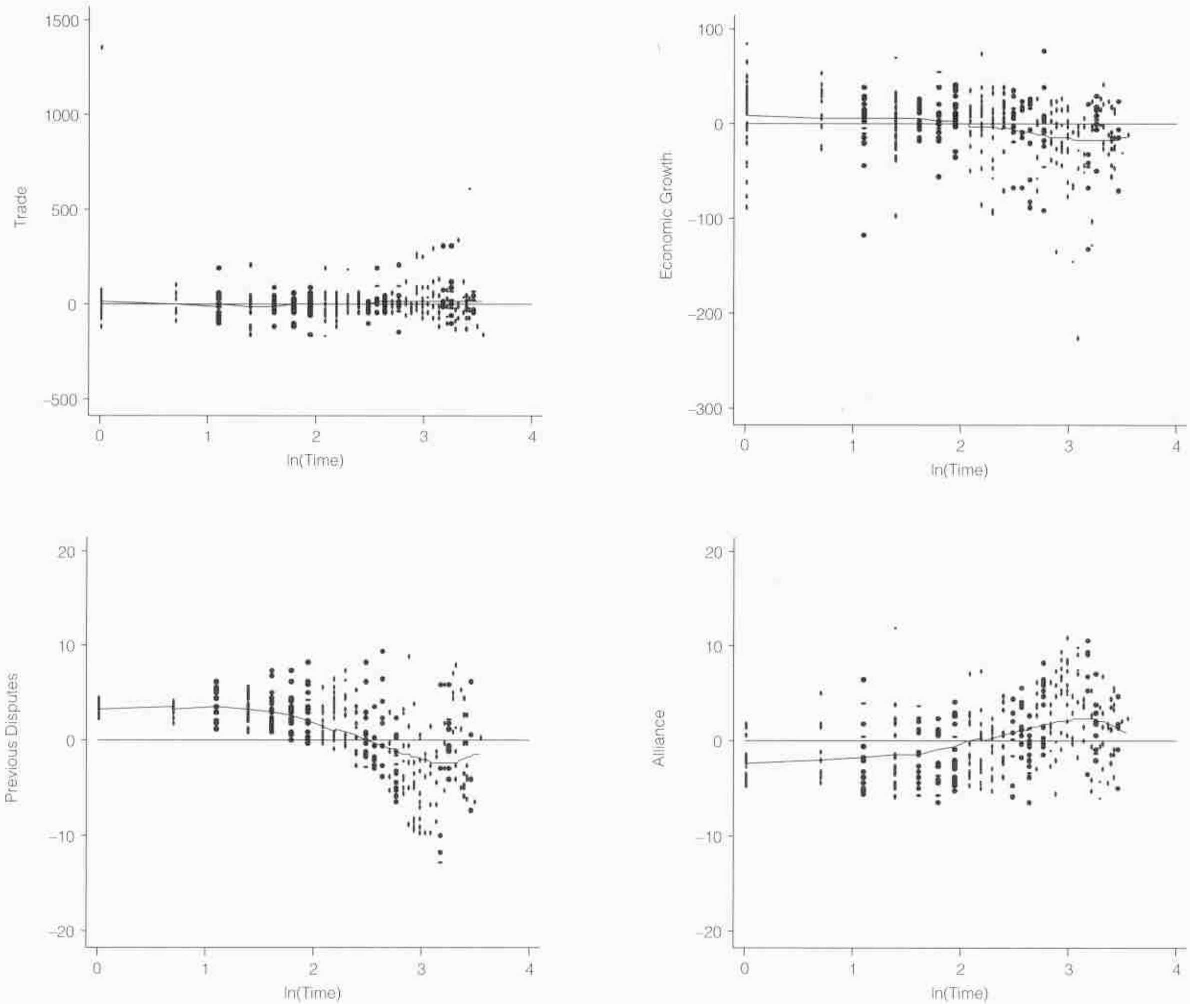
Taken as a whole, the evidence from the various tests for nonproportionality is clear. In a model of international conflict that fails to account for the effects of previous disputes, there are strong indications that the effects

FIGURE 2 **Plots of Scaled Schoenfeld Residuals against *ln*(Time): Model Without Previous Disputes**



Note: Figures plot scaled Schoenfeld (1982) residuals against *ln*(Time) for each variable. Lines are lowess smooths (bandwidth = 0.8). See text for details.

FIGURE 3 Plots of Scaled Schoenfeld Residuals against *ln*(Time), Model With Previous Disputes (Selected Variables)



Note: Figures plot scaled Schoenfeld (1982) residuals against *ln*(Time) for selected variables. Lines are lowess smooths (bandwidth = 0.8). See text for details.

TABLE 2 Results of Grambsch and Therneau Nonproportionality Tests, International Dispute Models

	Model Without Previous Disputes			Model With Previous Disputes		
Variable	ρ	χ^2	*p*-value	ρ	χ^2	*p*-value
Democracy	0.122	14.85	<.001	−0.119	7.32	.007
Economic Growth	−0.118	7.78	.005	−0.251	45.02	<.001
Alliance	0.146	31.88	<.001	0.439	209.49	<.001
Contiguity	0.069	4.67	.03	−0.251	39.23	<.001
Capability Ratio	0.111	25.47	<.001	0.051	2.40	.121
Trade	−0.097	9.08	.003	0.031	0.21	.650
Previous Disputes	—	—	—	−0.529	713.78	<.001
Global Test	—	61.85	<.001	—	761.76	<.001

Note: Results are based on models presented in Table 1, and are for log-time specifications; see text for details.

of international alliances and capability differences on preventing disputes wane over time, while that for economic growth grows stronger. These differences persist, and are magnified, in a model that also controls for the incidence of past disputes between the countries in question. Moreover, the pacific effect of such disputes themselves also declines dramatically over time, a finding consistent with the theory that the disputatious effect of recent conflicts "wears off" the longer the nations in question remain at peace. In each of these cases, a model that assumes (and estimates) a proportional effect for these covariates provides at best an incomplete, and at worst a misleading, picture of the true nature of the dependence.

To address this nonproportionality, and to better assess the effects of the variables just discussed, we estimate two models of international disputes in which we interact each of the potentially nonproportional covariates with *ln*(*Time*).[16] Doing so allows each covariate's effect on the hazard of conflict to vary monotonically with the duration of the peace and will thus provide a more complete and accurate picture of the true influence of these variables on the hazard of a dispute.

The estimates of the interactive models, both including and excluding *Previous Disputes*, are presented in Table 3. Because of the variables' interactions with log-duration, estimates of the direct effects have a natural interpretation as the effect of that covariate on the hazard of a conflict in the first year following the end of a dispute (that is, when $T = 1$). Consider first the model omitting previous disputes. Consistent with earlier results presented in Table 1, we see that the pacifying effects of an alliance, strong in the initial years after a conflict, wane as the peace wears on, such that after twenty-seven years of peace its influence is effectively zero. We suspect that this change is due to the fact that alliances which occur shortly after a dispute reflect a particularly high degree of commitment to peace between those nations, while those which persist, or are formed later, may lack the same level of commitment.[17] In addition, as time passes following the formation of an alliance, foreign policy preferences may diverge. With such changes, the chances of conflict between allied states (who by definition have some common interests) will also be likely to increase over time. A similar dynamic holds for differences in military capabilities: the initially large, negative impact of capability differences on the hazard of conflict declines over time, albeit at a less dramatic rate than that for alliances. Conversely, the pacifying effects of economic growth become stronger as the peace wears on; that variable's influence, initially negligible, grows dramatically more significant over time.

The results for the model that includes previous disputes are broadly similar, albeit with a few important differences. The effects for disparities in capabilities are more modest in the second analysis, as are the increases, with time, in the effect of economic growth on reducing conflict. Chief among the cross-model differences is the influence of alliances, which now exhibit *no* significant negative effect on the probability of a conflict, irrespective of when the alliance takes place. Moreover, alliances' effects on conflict decline to zero after only three years, and after that becomes steadily more positive, suggesting that alliances may actually act to increase the odds of a conflict after many years of peace.[18]

A second important difference is in the effect of previous disputes. Beck, Katz, and Tucker's (1998) analysis finds a strong positive influence of prior disputes: within a dyad, conflict tends to beget itself, with greater numbers of previous disputes being associated with higher probabilities of conflict. Our result confirms and extends this finding; importantly, however, our work also shows that the influence of previous disputes declines over time. The interactive model allows us to estimate the extent to which such disputes' influence on future conflict wanes over time. Figure 4 plots the odds ratios (i.e., the exponentiated coefficient estimates) for the previous disputes variable, as a function of time.[19] The large, positive odds ratio is consistent with the sizeable direct effect estimate: in the early years following a conflict, the effect of previous disputes is to increase the hazard of another conflict by as much as a factor of forty. As time progresses, however, that effect decreases rapidly, such that by the tenth year its effect is to increase the odds of a further dispute by 350 percent, and in the twentieth year of peace, by only 131 percent. And at the maximum value of observed duration ($T = 35$), the estimated effect of a pre-

[16] Note that the *ln*(*Time*) interactions used with the Cox model are easily adaptable to a discrete time model as well (e.g., Beck, Katz, and Tucker 1998). That approach simply involves including *ln*(Time) and the appropriate time-by-covariate interactions among such a model's explanatory variables; estimation and interpretation may then proceed normally.

[17] Such a theory squares well with recent research on wars as part of a broader bargaining game between states (e.g., Wagner 2000; Werner 1998, 2000a).

[18] While we are hesitant to place any great substantive stock in this result, we do note that it is consistent with the notion that alliances, like trade, provide "points of contact" between states and thus may be indicative of greater interaction (and thus, greater potential for dispute) than between nonallied states.

[19] The odds ratio indicates the percentage change in the hazard associated with a one-unit increase in the covariate in question; thus, an odds ratio of 2.0 corresponds to a 100 percent increase in the hazard (Box-Steffensmeier and Zorn 2001).

TABLE 3 Cox Models with Log-Time Interactions for Nonproportionality

Variable	No Previous Disputes	With Previous Disputes
Democracy	−0.692**	−0.266
	(0.208)	(0.154)
Economic Growth	0.688	−0.293
	(2.452)	(1.528)
Alliance	−1.211**	−0.287
	(0.294)	(0.198)
Contiguity	0.859**	0.447*
	(0.290)	(0.193)
Capability Ratio	−0.432**	−0.047
	(0.186)	(0.028)
Trade	20.643	−2.085
	(21.470)	(5.990)
Previous Disputes	—	3.731**
		(0.174)
Democracy × *ln*(Time)	0.123	0.03
	(0.084)	(0.071)
Economic Growth × *ln*(Time)	−2.040*	−1.407
	(1.030)	(0.762)
Alliance × *ln*(Time)	0.368**	0.282**
	(0.118)	(0.109)
Contiguity × *ln*(Time)	0.167	−0.005
	(0.126)	(0.089)
Capability Ratio × *ln*(Time)	0.103*	
	(0.063)	—
Trade × *ln*(Time)	−14.894	
	(9.419)	—
Previous Disputes × *ln*(Time)	—	−0.967**
		(0.062)
*ln*L	−2489.97	−1923.81

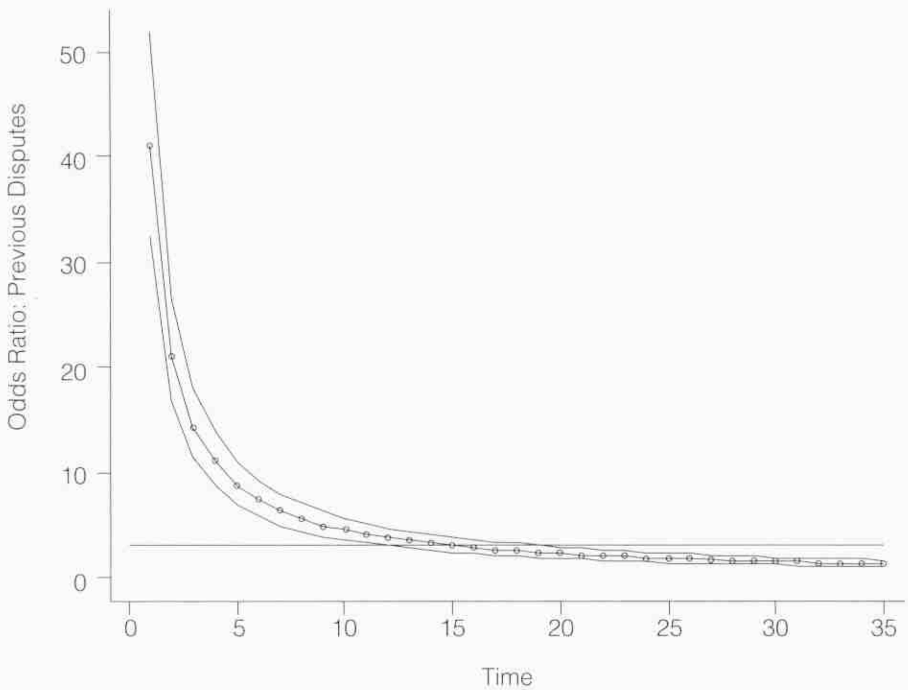
Note: Cell entries are estimated coefficients; robust standard errors are in parentheses. $N = 20{,}448$. One asterisk indicates $p < .05$, two indicate $p < .01$ (one-tailed). See text for details.

vious dispute is to increase the hazard of a conflict by a mere 35 percent. By contrast, the model that assumes a proportional effect for previous disputes (Table 1, column four) would suggest that the presence of a previous dispute raises the hazard of conflict by a uniform 189 percent, irrespective of how long ago that dispute occurred. This effect is indicated by the horizontal line in Figure 4; note that, for most of the durations in question, this estimate is outside of the 95 percent confidence interval for the actual, time-dependent effect. Figure 4 thus graphically illustrates how a model that assumes proportional hazards yields significantly misleading inferences about the influence of nonproportional covariates.

Conclusions

Duration models are fast becoming one of the most widely-used quantitative techniques in political science. As models of durations become increasingly common in our discipline, it is important that their properties and underlying assumptions be properly understood and appreciated. One such property is that of proportionality in covariate effects, and the validity of one's estimates depends strongly on meeting this assumption. Violations of this property have the potential for widespread and serious consequences for political scientists, for two reasons. First, as described and illustrated above, such violations

FIGURE 4 Odds Ratios by Duration: Previous Disputes Variable



Note: Figure plots the estimated odds ratio for the Previous Disputes variable, by duration, as estimated in Table 3. Circled line is estimated odds ratio; smooth lines are 95 percent confidence intervals, based on robust standard errors. See text for details.

can have dramatic and detrimental effects on parameter estimates, and therefore on the conclusions we draw about the processes under study. Second, proportionality in covariate effects is quite likely to be the exception rather than the rule. Thus, we are faced with a problem that is likely to be both serious and widespread in our research.

Mindful of this possibility, we have presented, discussed, and implemented a range of currently available tests for proportionality in duration models. We concentrated on the Cox model and presented both graphical and statistical approaches for detecting and ameliorating deviations from proportionality. Using data on international conflict, we demonstrated the importance of the proportionality property, showing how incorrect assumptions about proportionality can have a large impact on estimates of covariate effects and lead to misleading inferences about the process being studied. In general, we recommend the use of residual-based tests for determining whether proportional covariate effects are, in fact, a reasonably accurate description of one's data. Such tests, and the residuals upon which they are based, are increasingly easy to obtain in all commonly-used software packages for analyzing duration data. In addition, we show that relaxing the proportional hazards assumption allows one both to correct for nonproportionality in covariate effects and to evaluate the substantively interesting phenomena which give rise to nonproportionality in the first place. In summary, we recommend that applied researchers take the requirement of proportionality in these models seriously, and that diagnostic tests and remedial analyses we outline become standard practice for scholars analyzing political science duration data.

Manuscript submitted January 27, 2000.
Final manuscript received February 14, 2001.

Appendix A
Proportionality in the Weibull Model

The Weibull model, like nearly all other parametric models, is a proportional hazards model, and is the most commonly used duration model in political science (e.g., Bennett 1997; Werner 1998).[20] The Weibull model may be written as:

[20]The popularity of the Weibull over the Cox model in political science is somewhat surprising, given that the Cox model imposes many fewer restrictions than the Weibull. While both are models of proportional hazards, the Weibull model requires additional parametric assumptions about the hazard rate as well (Box-Steffensmeier and Jones 2001).

$$h(t) = \lambda p(\lambda t)^{p-1} \tag{A1}$$

where we typically specify $\lambda_i = exp(X_i\beta)$ and p is often referred to as the "shape" parameter. A shape parameter equal to one corresponds to constant hazards, while $p < 1$ indicates that hazards are decreasing and $p > 1$ suggests that hazards are rising over time.[21]

The Weibull model is a proportional hazards model because the ratio of the hazards for individuals i and j depends only on the covariates and p, and not on time:

$$\frac{h_i(t)}{h_j(t)} = \left(\frac{\lambda_i}{\lambda_j}\right)^p \tag{A2}$$

In the context of the Weibull model, in addition to covariate effects, the assumption of proportional hazards extends to the restriction that the shape parameter p is equal over time and across different values of the independent variables. That is, "in the Weibull model, the assumption of proportional hazards across a number of groups, g, say, corresponds to the assumption that the shape parameter in the baseline hazard function is the same in each group" (Collett 1994, 195; see also Zorn 2000). As in the case of the Cox model, if unaccounted for nonproportionality exists in the Weibull model, estimates of the influences of covariates are likely to be distorted.

The literature on testing the proportional hazards assumption is much less well developed for the Weibull than for the Cox model. In particular, many of the recently-devised residual-based tests for nonproportionality have yet to be applied in a parametric context. However, as pointed out earlier, one test available for testing proportionality in parametric models is to fit separate Weibull models to each of the g groups (Collett 1994). The values of the log-likelihoods lnL_g for each group can then be summed, and compared to that for the model combining all groups of data (and thus assuming common covariate effects, as well as a common shape parameter p). Twice the difference between these statistics follows a chi-squared distribution with $(g - 1) \times k$ degrees of freedom; failure to reject the null hypothesis can be taken as evidence that the assumption of proportional hazards is justified.

As a brief illustration, we estimated Weibull models of international conflicts, both with and without the *Previous Events* variable, using the data examined herein and dividing the data at $T \geq 14$.[22] For the model omitting previous disputes, the likelihood-ratio statistic is 42.58 (that is, $-2[-1029.64 - (-758.74 - 249.61)]$); for the model which includes *Previous Disputes*, the corresponding statistic is $-2[-709.56 - (-517.33 - 116.52)] = 151.42$. In both cases, we can clearly reject the joint null hypothesis that the covariates' effects on the hazards and the shape parameters are constant over time. As in the case of the Cox model, subsequent to such a finding, one can use time-by-covariate interactions to investigate the nature and extent of the nonproportionality. As noted above, the potential for collinearity among such interactive terms is high; overuse of such terms therefore runs the risk of decreasing the precision of one's estimates, and should be avoided.

Appendix B
Coding and Summary Statistics for International Dispute Data

Variable	Mean	Standard Deviation	Minimum	Maximum
Dependent Variable				
Duration	14.185	8.805	1	35
Independent Variables				
Democracy	−0.344	0.695	−1	1
Economic Growth	0.008	0.034	−0.265	0.165
Alliance	0.356	0.479	0	1
Geographical Contiguity	0.310	0.462	0	1
Capability Ratio	1.668	4.479	0.01	78.930
Intradyadic Trade	0.002	0.008	0	0.177
Previous Disputes	0.273	0.718	0	8

Note: N = 20,448 (827 dyads averaging 24.7 years per dyad and 405 total disputes). Data are taken from Oneal and Russett (1997) and Beck, Katz, and Tucker (1998).

[21]Many authors (e.g., Lancaster 1990) discuss the Weibull shape parameter in terms of σ, where $\sigma = 1/p$.

[22]For reasons of space, the results of these estimates are not presented here; they can be obtained from the authors upon request.

Variable Specifications

Duration	Duration of peace, in years.
Democracy	Polity III democracy score of the less-democratic dyad member, rescaled to range between –1 and 1.
Economic Growth	Annual GDP growth for the lower-growth dyad member, divided by 100.
Alliance	Coded 1 if the members of the dyad are allied, 0 otherwise.
Contiguity	Coded 1 if the dyad members are contiguous states, 0 otherwise.
Capability Ratio	The natural logarithm of the ratio of the two states' military capabilities, as measured by the Correlates of War data.
Trade	The ratio of bilateral trade to GDP, in constant U.S. dollars.
Previous Disputes	A running counter of the number of previous disputes experienced by the dyad during the 1950-1985 period.

References

Andersen, P.K., O. Borgan, R. D. Gill, and N. Keiding. 1993. *Statistical Models Based on Counting Processes.* New York: Springer-Verlag.

Andersen, P. K., and R. D. Gill. 1982. "Cox's Regression Model for Counting Processes: A Large Sample Study." *Annals of Statistics* 10:1100–1120.

Barbieri, Katherine. 1996. "Economic Interdependence: A Path to Peace or a Source of Conflict?" *Journal of Peace Research* 33:29–49.

Beck, Nathaniel. 1998. "Modeling Space and Time: The Event History Approach." In *Research Strategies in the Social Sciences,* ed. E. Scarbrough and E. Tannenbaum. New York: Oxford University Press.

Beck, Nathaniel. 1999. "Trade and Conflict in the Cold War Era: An Empirical Analysis." Presented at the Annual Meeting of the American Political Science Association, Atlanta, Ga.

Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42:1260–1288.

Bennett, D. Scott. 1997. "Testing Alternative Models of Alliance Duration, 1816–1984." *American Journal of Political Science* 41:846–878.

Box-Steffensmeier, Janet M., and Bradford S. Jones. 2001. *Timing and Political Change: Event History Models in Political Science.* Unpublished manuscript. Ohio State University.

Box-Steffensmeier, Janet M., Dan Reiter, and Christopher J. W. Zorn. 2000. "Substantive and Methodological Innovations for the Study of Temporal Dynamics in International Conflict." Presented at the Mershon Center Conference on Interdependence and Conflict, Columbus, Ohio.

Box-Steffensmeier, Janet M., and Christopher J. W. Zorn. 2001. "Duration Models for Repeated Events." Unpublished manuscript. Ohio State University.

Chastang, Claude. 1983. "A Simulation Study in a Two-Covariate Survival Model: Importance of the Proportional Hazards Assumption." *Controlled Clinical Trials* 4:148.

Chen, Chen-Hsin, and P. C. Wang. 1991. "Diagnostic Plots in Cox's Regression Model." *Biometrics* 47:841–850.

Chung, Ching-Fan, Peter Schmidt, and Ann D. Witte. 1991. "Survival Analysis: A Survey." *Journal of Quantitative Criminology* 7, 1:59–98.

Collett, Dave. 1994. *Modelling Survival Data in Medical Research.* London: Chapman & Hall.

Cox, D. 1972. "Regression Models and Life Tables." *Journal of the Royal Statistical Society, Serial B* 34:187–220.

Deshpande, Jayant V., and Debasis Sengupta. 1995. "Testing the Hypothesis of Proportional Hazards in Two Populations." *Biometrika* 82:251–261.

Diermeier, Daniel, and Randy T. Stevenson. 1999. "Cabinet Survival and Competing Risks." *American Journal of Political Science* 43:1051–1068.

Fleming, Thomas R., and David P. Harrington. 1991. *Counting Processes and Survival Analysis.* New York: Wiley.

Gartzke, Erik. 1998. "Kant We All Just Get Along?: Opportunity, Willingness and the Origins of the Democratic Peace." *American Journal of Political Science* 42:1–27.

Grambsch, Patricia M., and Terry M. Therneau. 1994. "Proportional Hazards Tests and Diagnostics Based on Weighted Residuals." *Biometrika* 81:515–526.

Grambsch, Patricia M., Terry M. Therneau, and Thomas R. Fleming. 1995. "Diagnostic Plots to Reveal Functional Form of Covariates in Multiplicative Intensity Models." *Biometrics* 51:1469–1482.

Gray, Todd Edward. 1996. "The Impact of Deviations From the Proportional Hazards Assumption on Power in the Analysis of Survival Data." Unpublished M.P.H. Thesis, Rollins School of Public Health, Emory University.

Guide to Statistics: S-Plus 2000 Modern Statistics and Advanced Graphics. 1999. Seattle, Wash.: MathSoft Data Analysis Products Division.

Harrell, F. 1986. *The PHGLM Procedure.* SAS Supplementary Library User's Guide, Version 5. Cary, N.C.: SAS Institute.

Holford, T. R. 1980. "The Analysis of Rates and Survivorship Using Log-Linear Models." *Biometrics* 32:299–306.

Hosmer, David W., Jr., and Stanley Lemeshow. 1999. *Applied Survival Analysis: Regression Modeling of Time to Event Data.* New York: Wiley.

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

Klein, John P., and Melvin L. Moeschberger. 1997. *Survival Analysis: Techniques for Censored and Truncated Data.* New York: Springer-Verlag.

Lagakos, S. W., and D. A Schoenfeld. 1984. "Properties of Proportional-Hazards Score Tests Under Misspecified Regression Models." *Biometrics* 40:1037–1048.

Laird, Nan, and Donald Oliver. 1981. "Covariance Analysis of Censored Survival Data Using Log-Linear Analysis Techniques." *Journal of the American Statistical Association* 96:231–240.

Lancaster, Tony. 1990. *The Econometric Analysis of Transition Data.* New York: Cambridge University Press.

Larsen, Ulla, and James W. Vaupel. 1993. "Hutterite Fecundability by Age and Parity: Strategies for Frailty Modeling of Event Histories." *Demography* 30:81–102.

Mansfield, Edward D., and Brian M. Pollins. 2000. "Economic Interdependence and Political Conflict: A Conceptual and Empirical Overview." Prepared for the Mershon Center Conference on Interdependence and Conflict, Columbus, Ohio.

Maoz, Zeev, and Bruce Russett. 1992. "Alliance, Contiguity, Wealth, and Political Stability: Is the Lack of Conflict Among Democracies a Statistical Artifact?" *International Interactions* 17:245–267.

Maoz, Zeev, and Bruce Russett. 1993. "Normative and Structural Causes of Democratic Peace, 1946–1986." *American Political Science Review* 87:624–638.

Martin, Lanny. 2000. "Public Opinion Shocks and Government Termination." Presented at the annual meeting of the Midwest Political Science Association.

McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models.* 2nd ed. London: Chapman and Hall.

Ng'andu, N. H. 1997. "An Empirical Comparison of Statistical Tests for Assessing the Proportional Hazards Assumption of Cox's Model." *Statistics in Medicine* 16:611–626.

Oneal, John R., Frances H. Oneal, Zeev Maoz, and Bruce Russett. 1996. "The Liberal Peace: Interdependence, Democracy, and International Conflict." *Journal of Peace Research* 33:11–28.

Oneal, John R., and Bruce Russett. 1997. "The Classical Liberals Were Right: Democracy, Interdependence, and Conflict, 1950–1985." *International Studies Quarterly* 41:267–294.

Oneal, John R., and Bruce Russett. 1999. "Assessing the Liberal Peace with Alternative Specifications: Trade Still Reduces Conflict." *Journal of Peace Research* 36:423–442.

Quantin, C., T. Moreau, B. Asselain, J. Maccario, and J. Lellouch. 1996. "A Regression Survival Model for Testing the Proportional Hazards Hypothesis." *Biometrics* 52:874–885.

Ray, James Lee. 1997. "The Democratic Path to Peace." *Journal of Democracy* 8:49–64.

Reed, William. 2000. "A Unified Statistical Model of Conflict Onset and Escalation." *American Journal of Political Science* 44:84–93.

Reiter, Dan. 1996. *Crucible of Beliefs: Learning, Alliances, and World Wars.* Ithaca, N.Y.: Cornell University Press.

Russett, Bruce. 1990. "Economic Decline, Electoral Pressure, and the Initiation of International Conflict." In *The Prisoners of War*, ed. Charles Gochman and Ned Allan Sabrosky. Lexington, Mass.: Lexington Book.

Russett, Bruce. 1993. *Grasping the Democratic Peace: Principles for a Post Cold War World.* Princeton: Princeton University Press.

Schemper, Michael. 1992. "Cox Analysis of Survival Data with Non-Proportional Hazard Functions." *The Statistician* 41:455–465.

Schoenfeld, David. 1982. "Partial Residuals for the Proportional Hazards Regression Model." *Biometrika* 69:239–241.

Singer, Judith D., and John B. Willett. 1993. "It's About Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events." *Journal of Educational Statistics* 18:155–195.

Teachman, Jay D., and Mark D. Hayward. 1993. "Interpreting Hazard Rate Models." *Sociological Methods and Research* 21:340–371.

Therneau, Terry M., Patricia M. Grambsch, and Thomas R. Fleming. 1990. "Martingale-Based Residuals for Survival Models." *Biometrika* 77:147–160.

Vermunt, Jeroen K. 1997. *Log-Linear Models for Event Histories.* Thousand Oaks, Calif.: Sage Publications.

Wagner, R. Harrison. 2000. "Bargaining and War." *American Journal of Political Science* 44:469–484.

Werner, Suzanne. 1998. "Negotiating the Terms of Settlement: War Aims and Bargaining Leverage." *Journal of Conflict Resolution* 42:321–343.

Werner, Suzanne. 2000a. "Deterring Intervention: The Stakes of War and Third Party Involvement." *American Journal of Political Science* 44:720–732.

Werner, Suzanne. 2000b. "The Effects of Political Similarity on the Onset of Militarized Disputes, 1816–1985." *Political Research Quarterly* 53:343–374.

Zorn, Christopher J. W. 2000. "Modeling Duration Dependence." *Political Analysis* 8:367–380.