ISSUES ACCOMPANYING TIME-VARYING DATA IN DURATION MODELING, WITH AN EMPIRICAL APPLICATION TO THE OVERRULING OF SUPREME COURT PRECEDENT

Brandon L. Bartels
Department of Political Science
Stony Brook University
Stony Brook, NY 11794-4392
brandon.bartels@stonybrook.edu

Abstract

Duration models with time-varying covariates (TVCs) present researchers with issues and opportunities of both the statistical and substantive varieties. In this paper, I highlight what I believe are the key issues related to time-varying data: (1) accounting for unobserved heterogeneity via frailty models and the split population duration model, and (2) cluster confounding and the estimation of within-subject and between-subject effects of TVCs. I also revisit the debate about parametric versus semi-parametric approaches and its relevance to time-varying data. To illustrate the methods and procedures highlighted, I reanalyze Hansford and Spriggs's (2006) data on the overruling of Supreme Court precedents. Models for dealing with unobserved heterogeneity—unshared and shared frailty models and the split population model—produce results that confirm some of Hansford and Spriggs's results yet refute others. Moreover, estimation of within and between-precedent effects of two of the authors' central theoretical variables—ideological distance and vitality—offers a compelling explanation regarding the precise nature in which these variables influence the survival of precedent. In sum, taking advantage of opportunities accompanying TVCs can further enhance one's understanding of the political process under examination.

Prepared for presentation at the 2007 Annual Meeting of the American Political Science Association, Chicago, IL. I thank Tom Hansford and Jim Spriggs for sharing their data with me.

Duration, or event history, models with time-varying covariates present both obstacles and unique modeling opportunities of both the substantive and statistical variety. The primary motivation for duration models is to model the timing of an event as a function of certain independent variables. What factors influence *when* an event will occur? In this paper, I raise several important issues and opportunities that accompany duration models with time-varying covariates that can have consequences for the types of inferences that are made. I also offer practical suggestions for analysts on these issues. The central topics include: (1) accounting for and modeling unobserved heterogeneity via frailty models and the split population duration model, (2) the continuing debate over parametric or semi-parametric models and choosing the appropriate distribution for parametric models, and (3) the issue of cluster confounding for TVCs and methods for estimating within-subject and between-subject effects for time-varying covariates. As I discuss below, not addressing or accounting for some of these issues can lead to invalid inferences about parameters of interest. Moreover, employing some of the methods I discuss can lead to enhanced and new substantive insights about duration processes.

To illustrate the points central to the paper, I reanalyze data from Hansford and Spriggs (2006, chapter 5; Spriggs and Hansford 2001) on the overruling of Supreme Court precedent. In particular, I show how different modeling specifications and operationalizations related to the topics mentioned above lead to substantively different inferences about the factors that influence when the Court will overturn its precedents.

DURATION MODELS AND TIME-VARYING COVARIATES

As is now fairly familiar in political science, duration data is time-to-event data, where the observed dependent variable is the amount of time (measured in, e.g., days, years, etc.) until some event occurs (for a general overview, see Box-Steffensmeier and Jones 1997, 2004). The theoretical dependent variable is the hazard rate, or the risk of the event occurring, given that it has not already occurred. The hazard of an event is then modeled as a function of covariates to analyze what factors influence the timing of an event's occurrence. Scholars have used duration models to examine the timing of cabinet failures (King et al. 1990), how long war lasts (Bennett and Stamm 1996), international alliances termination (Bennett 1997), position-taking by members of Congress (Box-Steffensmeier et al. 1997; Caldeira and Zorn 2004), when war will break out between two countries (Beck et al. 1998), and when the Supreme Court will overturn its own precedents (Hansford and Spriggs 2006; Spriggs and Hansford 2001). For some of these analyses (e.g., King et al. 1990; Box-Steffensmeier et al. 1997), data are single-spell data and resemble cross-sectional data; the units of analysis are subjects, countries, or entities, and the independent variables are unit specific and do not change over analysis time.

For other types of duration data (e.g., Bennett 1997; Beck et al. 1998; Hansford and Spriggs 2006), multiple spells are observed for each subject due to the inclusion of time-varying covariates (hereinafter, TVCs). Such data resemble panel data. That is, the values of some of the independent variables change over analysis time, allowing one to assess how changing conditions or circumstances influence the hazard of an event occurring. To use multilevel data terminology, TVCs are akin to level-1 variables (at the "spell" level) and time-constant independent variables are akin to level-2 variables (at the subject level). The nesting structure is spells, or occasions (level-1 units), nested within subjects (level-2 variables). Table 1 presents a simple example of this logic. The table presents 20 total spells and 5 subjects. In this example, imagine that each spell represents a year, meaning that analysis time is measured in years.[1] The "Event" column is the censoring indicator; "0" means that the subject has not yet experienced the

[1] Note that this example more closely resembles discrete-time data than continuous-time data. It is meant to be illustrative and simple. My focus in this paper is on continuous-time duration data.

event for a given spell, and “1” indicates that the subject has experienced the event. Note that subjects 1, 4, and 5 never experienced the event and are therefore censored subjects. Subject 2 experienced the event in 6 years, while subject 3 experienced the event in 4 years. There are three independent variables: X1, X2, and X3. X1 is a TVC, since for each subject, the values of the variable change over analysis time. X2 and X3 are both time-constant covariates. X2 is a subject-specific dummy variable, and X3 is also a subject-specific variable. In multilevel terms, X1 can be considered a level-1 variable since it varies across level-1 units (spells), while X2 and X3 can be considered level-2 variables since they only vary between subjects, but not between spells within subjects. For more examples and discussions of spells and time-varying data, see Bennett (1999), Box-Steffensmeier and Jones (1997, 2004), and Cleves et al. (2004).

[Table 1 about here]

Aside from time-varying data like the example in Table 1, other types of multilevel duration data are possible where some other nesting structure occurs. For instance, if one were studying the timing of voter decision making in an election (e.g., Box-Steffensmeier and Kimball 1999), one might be interested in nesting individual voters within congressional districts or states to account for unobserved heterogeneity across these higher-level, contextual units of analysis. While the focus in this paper is on time-varying data and the example in Table 1, but not on more general multilevel topics, the insights provided here can easily be applied to multilevel data like the timing of voting example. Below, I discuss the three major issues central to the paper and to duration data with TVCs: (1) heterogeneity, (2) issues relating to parametric models, (3) the issue of cluster confounding regarding TVCs and methods to correct for it.

TAKING HETEROGENEITY SERIOUSLY WITH TIME-VARYING DATA

One of the primary issues with duration data, particularly data with time varying covariates or a multilevel structure, is unobserved heterogeneity, or the notion that observations have different propensities of experiencing the event that cannot be accounted for by observed independent variables. For duration data with both censored and uncensored observations, there is most likely a large degree of heterogeneity in the propensity to experience the event. Standard duration models (like the Cox or Weibull models) assume that, conditional on observed covariates, observations are homogeneous, that is, they have the same propensity to experience the event of interest. Going a step further, standard models assume that all observations will eventually experience the event. Though one can account for observed variables that may account for the heterogeneity present in the data, there are often unobserved or unimagined factors that cannot be included in a model. Not accounting for this unobserved heterogeneity in some way can lead to incorrect estimates of coefficients and incorrect inferences about duration dependence for parametric models. As Box-Steffensmeier and Jones (2004, Chapter 9) note, two methods for dealing with heterogeneity include frailty models and the split-population duration model. I discuss both in relation to time-varying data below.

Individual and Shared Frailty Models

Frailty models are essentially random effects duration models that account for unobserved heterogeneity. Vaupel et al. (1979) introduced the term "frailty" to mean that some subjects are more prone to experience the event (i.e., some are more "frail") than others for reasons that cannot be accounted for by observed independent variables (Box-Steffensmeier and Jones 2004; Hougaard 2000). There are two types of frailty specifications: the individual (or unshared) frailty and the shared frailty. An individual frailty specification can be used for simple, cross-sectional duration data with time-constant covariates. I will discuss shortly how the individual frailty applies to duration data with TVCs. To see how the individual frailty enters into a typical duration model, consider first the hazard function for a basic, non-frailty model:

(1) $$h(t \mid x_j) = h_0(t)\exp(x_j'\beta)$$

The model is written in proportional hazard format.[2] The hazard rate is a function of a baseline hazard, $h_0(t)$, and covariates, x_j (a vector of independent variables) for j observation in the data. The vector β represents the effects of the independent variables on the hazard of the event occurring. In parametric models, the baseline hazard takes on a specific distributional form (e.g., an exponential or Weibull distribution), while the Cox model leaves the baseline hazard unparameterized.

To specify an individual frailty, the frailty term enters the hazard multiplicatively:

(2) $$h(t \mid x_j, \nu_j) = h_0(t)\nu_j \exp(x_j'\beta)$$

The frailty term, ν_j, represents unobserved heterogeneity across the j observations; it accounts for unobserved reasons why some observations might be more or less prone to experience the event than others. In a model using time-varying data with an individual frailty term, the observations would be *spells* as opposed to subjects. ν_j is always a positive quantity, which is why a gamma distribution is most often applied to ν_j.[3] For identification, ν_j is assumed to have a mean of 1 and an estimable variance, θ. If θ=0, the frailty model reduces to a standard, non-frailty model.[4]

A shared frailty model is akin to a random effects (or random intercept) model in panel or multilevel data. Data are clustered (or nested), and one accounts for unobserved heterogeneity

[2] The model can also be written in the accelerated failure time (AFT) metric, where the model is written in terms of the natural log of analysis time, ln(t).
[3] An inverse-Gaussian distribution can also be applied to the frailty term.
[4] Technical details about deriving the survival function and the likelihood function are given in Box-Steffensmeier and Jones (2004, Chapter 9) and Hougaard (2000).

across the higher-level units of analysis and dependence within clusters. In the case of TVCs, the nesting structure is spells nested within subjects, so a shared frailty is a time-constant, subject-specific term that represents each subject having a different propensity to experience the event, conditional on observed covariates. The hazard for the shared frailty model is written as:

$$(3) \qquad h(t \mid x_{ij}, \nu_j) = h_0(t)\nu_j \exp(x_{ij}'\beta)$$

where *i* indexes spells, and *j* indexes subjects.

For time-varying data, an important substantive and statistical distinction exists between the unshared and shared frailty models. The unshared frailty model in expression 2 does not account for the nesting structure of spells nested within subjects. Thus, the frailty term accounts for unobserved heterogeneity across *spells* as opposed to across subjects; the frailty accounts for each spell having a different propensity of experiencing the event due to unobserved factors. The shared frailty model accounts for unobserved heterogeneity across subjects as opposed to across spells. In the shared frailty model, subjects have time-constant propensities to experience the event due to unobserved factors. Analysts should carefully consider which type of frailty specification is most important for the research question at hand. In some cases, accounting for spell-specific heterogeneity might be most appropriate, but in many cases, it is often more appropriate to account for subject-specific heterogeneity. In his analysis of alliance duration, Bennett (1997) specifies a spell-specific (individual) frailty, suggesting that there is unobserved heterogeneity across spells. But Bennett does not specify a shared frailty model that would suggest that *alliances* (as opposed to spells within alliances) have different propensities to experience the event due to unobserved factors. Aside from theoretical considerations, one can also use information criterion measures, such as the Akaike information criterion (AIC) model statistic, as a guide to choosing between the unshared and shared frailty models.

The Split Population Duration Model

Standard duration models assume that every observation will eventually experience the event of interest. This strong assumption often poses a potential violation of the process under examination, particularly for data that contain a significant number of right-censored observations. Heterogeneity exists, then, when there is reason to believe that not all observations will experience the event. The split population (SP) model accounts for a specific type of heterogeneity, i.e., the possibility that some cases will never experience the event of interest while some will. It does so by modeling both *whether and when* an event will occur. As detailed in Schmidt and Witte (1984, 1989), the SP model essentially "splits" the observations under analysis into two subpopulations: one that will eventually experience the event of interest and one that will never experience the event. In their study of criminal recidivism, Schmidt and Witte (1989) specified a model that generates two sets of simultaneously estimated coefficients: one for the effects of covariates on whether the event will occur and another for the effects on covariates on the timing of the event, conditional on the event ever occurring. The mathematical derivation of the model is contained in Schmidt and Witte (1989), Box-Steffensmeier and Jones (2004), and Box-Steffensmeier et al. (2005).

It is worth emphasizing a few important points about the SPD model. In the SP model, the censoring indicator (i.e., whether or not we observe the event occur within the analysis time) serves as the dependent variable in the incidence portion of the model. Second, a very powerful feature of these models is that different covariates can be included to explain whether and when the event occurred. For example, an independent variable may have a positive effect on whether the event occurred and a negative effect on when it occurred. Third, the SPD model estimates a "split parameter," δ, which is the estimated mean probability experiencing the event of interest.

This statistic allows one to test whether relaxing the assumption that every observation will experience the event of interest is necessary. If it is not, i.e., if $\delta = 1$, the SP model collapses to a typical duration model. The estimated split also serves as a sort of goodness-of-fit statistic in that it can be compared to the proportion of cases that actually experienced the event of interest. Fourth, as far as I know, SP models are currently only estimable using parametric approaches. For those who advocate the semi-parametric approach to duration modeling, this is certainly the downside of the SP model. However, work continues to be done to estimate a Cox-type SP model, although problems of model identification have hampered these efforts (see, e.g., Sy and Taylor 2000; Kuk and Chen 1992).

In political science, scholars have yet to apply the SP model to duration data with TVCs. There have been SP applications to single-spell data. Hettinger and Zorn (2005) apply the SP model to examine both whether and when Congress will override a Supreme Court decision; this study correctly relaxes the assumption that all Court decisions will eventually be overturned. Box-Steffensmeier, Radcliffe, and Bartels (2005), recognizing that not all labor and corporate PACs will not eventually give donations to every member of Congress, analyze the factors that influence the incidence and timing of PAC contributions. Finally, in the age of contentious nomination processes to the U.S. Court of Appeals where not all nominees are eventually confirmed, Scherer, Bartels, and Steigerwalt (2007) apply an SP model to test how interest group opposition (and other factors) influences both the incidence and timing of confirmation.

While the SP model for TVCs has not been applied in political science, in the statistics literature, Forster and Jones (2001) have developed and applied an SP model with TVCs to explain factors the influence whether and when a person starts smoking. The model holds great promise for studies with time-varying data and the need to relax the assumption that all observations will experience the event of interest. In the empirical section, I show how the SP can be applied to time-varying data on the overruling of Supreme Court precedent.

PARAMETRIC AND SEMI-PARAMETRIC MODELS REVISITED

Debates about whether parametric or semi-parametric models should be employed have been hashed out extensively (e.g., Box-Steffensmeier and Jones 1997, 2004; Bennett 1999). The primary difference is how each approach treats duration dependence, which is defined as whether the hazard, conditional on covariates, is significantly influenced by the amount of time previously spent without experiencing the event. For example, in a Weibull model, positive duration dependence means that the conditional hazard increases over analysis time, meaning that the more time spent in a state, the greater the risk of experiencing the event. Bennett (1999) argues that duration dependence is a significant feature of duration models that should be interpreted substantively, i.e., ascertaining whether certain processes become entrenched or institutionalized over time. In a Weibull model, negative duration dependence suggests such an institutionalization process. In Bennett's (1997) alliances study, institutionalization would mean that the more time spent in the alliance, the lower the risk of that alliance terminating.[5] On the other hand, the Cox model treats duration dependence as a nuisance to be controlled for, as opposed to a key substantive feature of the model.

Choice of the parametric versus the semi-parametric approach rests on the goals of the research endeavor. If one is not substantively interested in duration dependence and is solely interested in the effects of covariates on the hazard, then the Cox model is appropriate. Choosing the Cox approach means not having to justify the choice of a parametric distribution for the baseline hazard. But choosing a parametric approach offers the analyst significant substantive

[5] Bennett (1997) actually finds positive duration dependence, a finding directly contradicting the institutionalization hypothesis.

insight and opportunities to communicate results in a way that is more intuitive to readers. In addition to the substantive appeal of making inferences about duration dependence and institutionalization, another significant advantage of the parametric approach is the ability to present post-estimation results communicating how covariates have an impact on the expected amount of time (e.g., days, years, etc.) until the event will occur. Analysts estimating Cox models often present post-estimation results in terms of the impact of a covariate on the percentage change in the hazard rate. A garden variety example of this procedure is provided in Binder and Maltzman's (2002, 196) study of the timing of a nominee's confirmation to the U.S. Courts of Appeals. Binder and Maltzman (2002, 196) communicate the following post-estimation interpretation for the effects of divided government: "the hazard rate of a [confirmation] decision decreases during divided government by nearly fifty percent compared to periods of unified control…." This is indeed a correct interpretation from a proportional hazards Cox model, but in many ways, the technique of presenting effects in terms of the proportional change in the hazard rate lacks substantive punch. A more powerful presentation would be to calculate the *expected number of days until confirmation* for both unified and divided government, while holding the remaining variables at some baseline value. Parametric models allows for such a presentation, while the Cox model does not.

In the parametric approach, a common way to communicate expected durations is to calculate the median duration time for a particular covariate profile. As discussed in Box-Steffensmeier and Jones (2004, 26), the median duration time for a Weibull model can be calculated as:

(4) $(\exp(-x'\beta))^{-1}(\log(2))^{1/p}$.

Equations for calculating median duration times for other parametric distributions are in Box-Steffensmeier and Jones (2004, Chapter 3).

Two more points are worth making regarding the use of parametric models. First, one must make an important choice about the distribution of the baseline hazard. Making this choice involves a mix of *a priori* expectations and statistical measures of fit. Analysts must first ask what type of distribution best matches the nature of the duration process under examination. If one has theoretical reasons to believe that the hazard of an event increases or decreases monotonically over time, then the Weibull is appropriate. If one believes there is no duration dependence and the hazard is constant over time, then the exponential is appropriate. The Weibull nests the exponential, so when p=1 (indicating no duration dependence) in the Weibull model, it reduces to the exponential. Other distributions, in the accelerated failure time (AFT) metric, are the log-logistic and log-normal distributions, which allow for non-monotonic hazards. If one expects the risk of an event to increase rapidly early in the analysis time and then taper off gradually as time passes, then the log-logistic or log-normal distribution is appropriate. For example, Hettinger and Zorn (2005) possessed this expectation with respect to congressional overrides of Supreme Court statutory decisions. Since the log-normal and log-logistic are not nested within the Weibull or exponential (or vice-versa), informational criterion measures such as the Akaike information criterion (AIC) can be used to compare non-nested models (see Box-Steffensmeier and Jones 2004, 44-45). A lower AIC value indicates a better fit, and the measure can be used to facilitate the choice of the most appropriate distribution, as well as the most appropriate model specification in general, e.g., in choosing between an unshared versus shared frailty specification.

The second point regarding parametric models relates to duration dependence in parametric models. As others (Bennett 1997, 1999; Box-Steffensmeier and Jones 1997, 2004; Zorn 2000) have noted, duration dependence, or more generally can change depending on what covariates are added to or subtracted from a model. Thus, analysts interested in making substantive conclusions about duration dependence and institutionalization must be confident that they have specified the most appropriate model. This brings to mind the distinction between “true” and “spurious” duration dependence, a distinction raised by, among others, Blossfeld and Rowher (2002), Heckman (1991), and Zorn (2000). The distinction is directly tied to the importance of accounting for unobserved heterogeneity via a frailty specification. Imagine a scenario where there is no duration dependence and significant heterogeneity between subjects. If one does not account for this heterogeneity, it will appear as if there is significant duration dependence because the heterogeneity will be fed into the baseline hazard (see Zorn 2000, 368-71). When one does account for unobserved heterogeneity, however, one can estimate true duration dependence. Thus, using a frailty specification in conjunction with a parametric approach can lead to substantive opportunities about communicating true duration dependence.[6]

CLUSTER CONFOUNDING WITH TVCs

A topic that has not received due attention regarding duration models with TVCs is the issue of cluster confounding. Cluster confounding occurs when data have multiple levels of analysis, as time-varying data in duration models do, and therefore, the effect of a variable is manifested at multiple levels. In the case of time-varying data, at level-1 (the spell-level), the effect of a TVC can be considered a *within-subject*, or *longitudinal, effect*, while at level 2 (the subject level), the effect of a TVC can be considered a *between-subject*, or *cross-sectional, effect*. This distinction is important in communicating the precise manner in which a TVC exhibits an

[6] Moreover, Zorn (2000) shows how one can model this true duration dependence as a function of covariates.

impact on the risk of event occurring. Does the effect represent how variation in the TVC *between subjects* influences the risk of event occurring? Or does the effect represent how variation over time, *for a given subject* (or, *within subjects*), in a TVC influences the hazard? Duration modelers rarely, if ever, separate TVCs into their within and between components (see, though, Zorn 2001) in a manner analogous to how multilevel modelers have suggested (e.g., Skrondal and Rabe-Hesketh 2004, 52-53; Bafumi and Gelman 2006).[7] By including the TVC as is, one makes the assumption that the within and between-subject effects are equal. As a result of not distinguishing within and between variation, one runs the risk of cluster confounding, which means that the within and between components of the variable are combined (or confounded) together, and one is not be able to distinguish the degree to which the effect is a longitudinal or cross-sectional one.[8] If the between and within effects of a TVC are the same (something we can test for), then cluster confounding is not a problem; the coefficient we estimate by including the originally-coded TVC will be statistically indistinguishable from the within and between effects of the TVC.[9] If we wanted to estimate the within and between effects of a TVC in the same model, one simply executes the following steps (see Skrondal and Rabe-Hesketh 2004, 52-53):

1. For a TVC, x_{ij}, calculate the mean value within each subject, i.e., the *subject mean*: $\bar{x}_j$.[10]
2. Calculate the *subject mean-centered* value of the variable by subtracting the

[7] This issue is discussed in the multilevel modeling context, e.g., Skrondal and Rabe-Hesketh (2004) and Bafumi and Gelman (2006). However, applied work in political science has yet to execute these procedures. I have yet to see the procedure discussed in the context of continuous-time duration models with TVCs, though Zorn (2001) has analyzed the issue with respect to discrete-time duration models, or what Beck, Katz, and Tucker (1998) call "binary TSCS" data.
[8] In a non-duration modeling context, effects from a random effects (aka, random intercept) model represent "partially-pooled" effects, that is, a weighted average of the within and between effects, depending on the degree of partial pooling in the data (see, e.g., Skrondal and Rabe-Hesketh 2004).
[9] Note how this discussion is analogous to the Hausman test in panel and TSCS data, which tests the equality of coefficients from the within (i.e., fixed effects) estimator and the random effects (partially-pooled) estimator.
[10] In the multilevel or panel data context, this would be the same as calculating the cluster means.

subject mean from the original value of the TVC: $x_{ij} - \bar{x}_j$.

The *subject mean* variable ($\bar{x}_j$), a time-constant variable, captures between-subject (or cross-sectional) variation in the TVC; it represents the average level of the TVC over analysis time. The *subject mean-centered* ($x_{ij} - \bar{x}_j$) variable, a TVC, captures within-subject (or longitudinal) variation; it represents change in the TVC over time for a given subject. Entering both of these variables into the duration model then allows one to estimates the separate between and within-subject effects of a TVC. The coefficient associated with $\bar{x}_j$ is the between-subject effect, while the coefficient associated with $x_{ij} - \bar{x}_j$ is the within-subject, or longitudinal, effect. Since the within and between operationalizations of the TVC completely separate within and between variation in the TVC, the correlation between $\bar{x}_j$ and ($x_{ij} - \bar{x}_j$) equals zero. One can then use a Wald test for the equality of the within and between effects.

An alternative procedure for testing the equality of the within and between-subject effects is to include the originally-coded TVC, x_{ij}, and the subject mean, $\bar{x}_j$ in the model. The within-subject effect will be the coefficient associated with x_{ij} (which will be equivalent to the within-subject effect from the previously discussed procedure). However, including x_{ij} (instead of $x_{ij} - \bar{x}_j$) changes the interpretation of the coefficient associated with $\bar{x}_j$. Now, the coefficient associated with $\bar{x}_j$ represents the *difference* between the within and between effects of the TVC. One can use a z-test for the coefficient to test whether the difference between the within and between effects is statistically significant (for more details, see Skrondal and Rabe-Hesketh 2004, 52-53).[11]

[11] Note that this second procedure is similar to what Bafumi and Gelman (2006) advocate in the multilevel modeling context, though they are not specific about what exactly the effects represent.

EMPIRICAL ANALYSIS: THE OVERRULING OF SUPREME COURT PRECEDENT

To illustrate the methods and procedures discussed above, I reanalyze Hansford and Spriggs's (2006, Chapter 5; Spriggs and Hansford 2001) data on the overruling of Supreme Court precedent. Hansford and Spriggs (2006), who use updated data and a revised model from their previous work on the topic (Spriggs and Hansford 2001), analyze the factors that influence when the Supreme Court will overturn its own precedents. The authors gathered data on 6,363 precedents that the Court decided between 1946-2001. Analysis time is years, so the duration variable is the number of years until a precedent is overturned by the Court. Of the 6,363 precedents, only 107 (1.7%) are eventually overturned. Though the authors include a host of relevant covariates (discussed below), there is more than likely a significant degree of unobserved heterogeneity in the data. Theoretically, we know that there is relative stability in the law on the Supreme Court. Though many subscribe to the notion that justices' choices are based on ideological considerations (e.g., Segal and Spaeth 2002), the Court does not exhibit volatile behavior in terms of overturning its own precedents. Indeed, *stare decisis* ("let the decision stand") is a norm that constrains the justices from producing frequent wholesale overrides of its cases (e.g., Knight and Epstein 1996). Given this theoretical backdrop, one would expect that precedents have significantly different propensities of being overturned due to unobserved factors, which implicates the use of a frailty specification. Moreover, it is reasonable to assume that not all precedents will eventually be overturned, thus making the split population model applicable.

Hansford and Spriggs include 11 covariates in their model. The two central covariates key to testing Hansford and Spriggs's theory are TVCs: (1) ideological distance between the median member of the majority that handed down the precedent and the median member of the

Court at time *t*;[12] and (2) the vitality of a precedent, or "the extent to which [a precedent] maintain[s] legal authority…. Some precedents are more legally authoritative than others and thus have an enhanced ability to justify and legitimize the justices' policy choices" (Hansford and Spriggs 2006, 23). Vitality is measured as the number of prior positive interpretations the Court has given to a precedent minus the number of negative interpretations. The authors hypothesize that the greater the ideological distance between the precedent and the Court at time *t*, the greater the risk of a precedent being overruled. Important to the theory, the authors also posit an interaction effect between ideological distance and vitality. They posit a conditional impact of vitality on the hazard of an overruling. The hypothesis states that vitality's effect of lowering the risk of a precedent being overturned will be diminished as ideological distance between the precedent and the Court at time *t* increases. In other words, vitality only matters when the Court is ideologically congruent with the precedent. Regarding the interaction, the authors also hypothesize that increases in vitality will make the impact of ideological distance on the risk of overturning a precedent even stronger. Put another way, the ideological basis for overturning a precedent is strengthened as a precedent's vitality increases. An ideologically-divergent Court would be much more motivated to strike a vital precedent than a non-vital one. A key mechanism for this effect is the Supreme Court's motivation to reinforce its control over lower courts, which would be more likely to follow vital precedents than non-vital ones.

In addition to ideological distance and vitality, Hansford and Spriggs include 9 control variables. Two of them, whether the issue area implicated by the precedent is active on the Court's agenda and the total number of prior interpretations the Court has given the precedent, are TVCs. The remaining 7 independent variables are time-constant variables. These include

[12] The authors use issue-specific preference measures for the ideological placement of the precedent (based on the median member of the majority) and the ideological position of the Court at time *t* (see Hansford and Spriggs 2006, 82).

characteristics specific to the precedent when it was decided (number of concurring opinions, the voting margin, the breadth or complexity of the precedent, whether it was a constitutional case) and variables measuring the salience of the precedent (whether the case was covered on the front page of the *New York Times*, the number of amicus curiae briefs filed, and whether the opinion was a *per curiam* or full opinion). For more information on these variables, see Hansford and Spriggs (2006, 59-63).

The total number of spells in the data is 179,653, and the total number of precedents is 6,363. To use multilevel terminology, we have 179,653 spells nested within 6,363 precedents. Hansford and Spriggs employ a Cox model to test their hypotheses. They do not, however, take any steps to account for heterogeneity or cluster confounding. My empirical analysis proceeds in three steps. First, I compare Hansford and Spriggs's results to individual (unshared) and shared frailty models. Second, I compare the authors' results to a split population model. Note that the models in the first two steps include the variables as Hansford and Spriggs originally coded them. Third, I address the cluster confounding issue and discuss results from a model that distinguishes within and between-precedent effects of the two central theoretical variables, ideological distance and vitality.

Frailty Models

I first compare Hansford and Spriggs's Cox results to two parametric frailty models: the Weibull and log-logistic.[13] In all models, I use a gamma distribution for the frailty term. In addition, I specify the Weibull model in the accelerated failure time metric. Regarding the baseline hazard, *a priori*, the Weibull seems appropriate if one expects the hazard to increase or decrease monotonically over time. From an institutionalization standpoint, I would expect

[13] Results from the log-normal were very similar to those from the log-logistic, so I chose not to report the log-normal results. Choosing between these two distributions is analogous to choosing between logit and probit in binary choice models.

negative duration dependence or no duration dependence at all, suggesting the hazard of a precedent being overruled does not increase over time. In a Weibull, if there is no duration dependence, then the shape parameter, *p*, equals 1, and the Weibull reduces to the exponential. A log-logistic distribution would allow for a non-monotonic hazard, and I would expect a similar pattern as what Hettinger and Zorn (2005) posited in their study of congressional overrides of Court decisions. That is, the hazard should increase rapidly early on in analysis time, and then taper off and decrease as time passes by. In fact, Hansford and Spriggs present a lowess-smoothed plot of the baseline hazard from their Cox model that very much resembles a log-logistic distribution. Statistically, I will discuss measures of fit—including the AIC—that will allow me to assess which model fits the data the best.

As I stated, the parametric approach often has advantages over the semi-parametric approach. As discussed, duration dependence is of substantive interest: To what extent does the conditional hazard of a precedent being overturned change over analysis time as a function of the previous time spent in a state of survival? Can precedents become institutionalized over time, making them essentially invincible? Or do precedents face a greater risk of being overturned as time passes? When accounting for unobserved heterogeneity via a frailty specification, one can, of course, recover the degree of true duration dependence. Next, for computational and comparability reasons, the parametric approach makes sense due to the fact that, as far as I know, the split population duration model is only estimable via parametric approaches. Thus, a focus on parametric approaches maintains comparability between the frailty specifications and the split population model, which I will discuss in further detail below. Finally, estimating a Cox frailty model has proved to be computationally difficult. Thus far, I have found it difficult to get the model to converge, and I am continuing to work on this issue.

Table 2 presents results comparing Hansford and Spriggs's Cox results to individual (spell-specific) frailty models. Table 3 compares Hansford and Spriggs's results to shared (precedent-specific) frailty models. Note how the individual frailty models (in Table 2) account for unobserved heterogeneity across spells, while the shared frailty models (in Table 3) account for time-constant unobserved heterogeneity across precedents. In comparing the Cox results to the frailty models in Tables 2 and 3, note that since the frailty models are in the AFT metric, the signs will be the opposite of those from the Cox model. In the Cox, positive coefficients indicate how increases in a covariate increase the hazard of the event occurring, or bring about its occurrence earlier in analysis time. In the frailty models (in the AFT metric), positive coefficients mean increases in a covariate will delay the event from occurring, or decrease the risk of an overruling.

[Table 2 about here]

In terms of model fit and comparison for the individual frailty models in Table 2, we first see that there is statistically significant spell-specific heterogeneity for both models. In the Weibull, θ=13.432, and a likelihood-ratio test suggests that the frailty model provides a significant improvement in fit over a non-frailty model (chi-sq=21.33, p<.001). The same goes for the log-logistic model. For the Weibull, note that the shape parameter, p, is 1.086, which indicates *positive* duration dependence and evidence against the notion that precedents become institutionalized over time (which would suggest a $p < 1$). However, results from a hypothesis test, where the null is p=1, suggest that there is no significant duration dependence (z=.58, p=.56), and therefore, that the Weibull basically reduces to an exponential. Though no support exists for institutionalization, the finding of no true duration dependence suggests that precedents exhibit a significant degree of stability over time. We know that the risk of a precedent being overruled is low to begin with. If this risk remains fairly stable over time, this bodes well for the long-term prospects of a precedent's survival. For the log-logistic model, the gamma parameter is .951, which suggests a non-monotonic hazard in which the baseline hazard first increases, peaks, and then decreases over time. In terms of model comparison between the Weibull and the log-logistic, the AIC statistics are very similar, suggesting that one model does not necessarily fit the data better than the other. Moreover, the parameter estimates and significance levels are very similar between the two frailty models.

A couple of important differences exist between results from the Cox model and the individual frailty models in Table 2. The effects of ideological distance and vitality in the first two rows are conditional effects due to the interaction term in the third row. Hansford and Spriggs found that when vitality equals 0, increases in ideological distance significantly increased the hazard of an overruling. Both frailty models produce the same result. Note that vitality ranges from -7 to 17, with zero representing no vitality and also, the median value of the variable. A difference emerges between the Cox results and individual frailty models with respect to the conditional impact of vitality. Spriggs and Hansford found, in accord with one of their central hypotheses, that when ideological distance equals zero (i.e., when the Court "likes" a precedent on ideological grounds), vitality significantly decreases the risk of a precedent being overruled. In the frailty models, however, these effects are statistically insignificant. Another important difference is in the interaction effect. Hansford and Spriggs found evidence—via a significant ideological distance by vitality interaction—that increases in vitality significantly enhanced the impact of ideological distance on the risk of an overruling. That is, the ideological motivations for overruling a precedent significantly increase as vitality increases. In both of the frailty models, however, the interaction between ideological distance and vitality is statistically insignificant. Thus, the individual frailty models seem to refute two of the three hypotheses put forth by Hansford and Spriggs. Among control variables, a couple of differences emerge. Though insignificant in the Cox model, the number of concurring opinions is significant in the frailty models; precedents with more concurring opinions face a greater risk of being overruled. Also, while the amici filings variable is significant in the Cox model, it is statistically insignificant in both frailty models.

Turning now to the shared frailty models in Table 3, both frailty models suggest evidence of statistically significant precedent-specific heterogeneity (vis-à-vis the θ estimates and the likelihood ratio tests for presence of heterogeneity). These results suggest there is significant time-constant, precedent-specific unobserved heterogeneity in the data; some precedents are more "frail," or prone to being overruled, than others due to unobserved factors. The Weibull model has a shape parameter, *p*, of 1.198, but like the individual frailty model, this parameter is not significantly different from 1 (z=1.20, p=.23), which means that the Weibull is not significantly different from the exponential and there is not significant duration dependence. For the log-logistic shared frailty model, which seems to be the model that best fits the data, note that γ=.863, again suggesting a non-monotonic baseline hazard.

[Table 3 about here]

To assess whether the spell-specific frailty specifications (in Table 2) or the shared precedent-specific frailty models (in Table 3) offer a better fit of the data, we can again turn to the AIC statistics. Recall that lower AIC statistics indicate a better fit. Importantly, note that, for both the Weibull and log-logistic models, the AIC statistics from the shared frailty specifications are lower than those from the individual frailty models. This suggests that accounting for time-constant precedent-specific heterogeneity, via the shared frailty model, provides a better fit than accounting for spell-specific heterogeneity. In comparing the shared frailty models, the AIC statistics suggest that the log-logistic model offers a slightly better fit than the Weibull model. This makes sense given that Hansford and Spriggs's lowess-smoothed baseline hazard from the Cox model closely resembled a log-logistic distribution.

In comparing the shared frailty models in Table 3 to the Cox model, a number of important differences again stand out. Both frailty models show that, when vitality equals 0, as ideological distance increases between a precedent and the Court at time *t*, survival time decreases. Unlike in the individual, spell-specific frailty results, the shared frailty models provide a result similar to the Cox model for the conditional effect of vitality (when ideological distance is zero). In the Weibull model, when the Court "likes" a precedent ideologically, vitality significantly decreases the risk of an overruling at the alpha=.05 level. In the log-logistic model, the same effect exists, and it is close to being statistically significant at the .05 level (p=.06). Another important difference is in the interaction effect. In both of the shared frailty models, the interaction between ideological distance and vitality is statistically insignificant. Unlike the Cox results, the frailty models show that the ideological motivations for overruling a precedent do not significantly increase as vitality increases. Whereas the individual frailty models refuted two out of Hansford and Spriggs's three hypotheses, the shared frailty models refute just one. A couple of differences exist between the Cox and the shared frailty models for the control variables as well. The number of concurring opinions is again significant in the shared frailty models. Also, the amici filings variable is statistically insignificant in both frailty models, while it is significant in the Cox model.

Figure 1 presents the survival functions for four covariate profiles from the log-logistic shared frailty model: high ideological distance, high vitality; high distance, low vitality; low distance, high vitality; and low distance, low vitality. Low and high values of ideological distance are the 5th and 95th percentiles, respectively. Low and high values of vitality are two standard deviations below and above the mean, respectively. Control variables are set at their mean values. This figure is akin to Hansford and Spriggs's (2006, 89) Figure 5.3.[14] Figure 1 demonstrates that when distance and vitality are high, a precedents survival probability remains near 1 throughout analysis time. On the other hand, when distance is high and vitality is low, the survival rate decreases the most out of the four profiles, though it only dips down to about .9 toward the end of analysis time. The remaining two profiles demonstrate similar survival propensities.

[Figure 1 about here]

Split Population Model

I now estimate a split population duration model of the overruling of precedent. As discussed earlier, the model generates two sets of coefficients: one for the likelihood, or incidence, of a precedent being overruled and the other for the timing of an overruling, conditional on it being overruled. I specify a log-logistic distribution for the hazard and a probit splitting function for the incidence of an overruling. The results are presented in Table 4. For a split population model with time-varying covariates, Forster and Jones (2001) suggest including only the time-constant covariates in the incidence portion of the model. I have followed this practice, excluding the four TVCs from this part of the model. The duration part of the model includes all covariates.

[Table 4 about here]

[14] One major difference is that Hansford and Spriggs set control variables with positive coefficients one standard deviation above the mean and control variables with negative coefficients one standard deviation below the mean. The authors' intention is to present survival functions for precedents with an above average chance of eventually being overturned.

The results from the SP model bear some resemblance to the frailty models, particularly the individual frailty models. First, γ=.91, which suggests a similar shape of the baseline hazard to those from the log-logistic frailty models. The estimated split is .169, which is the average predicted probability of a precedent being overruled. However, the actual split is a mere .017. This rather large difference between the estimated and actual splits suggests that the model's goodness-of-fit is not ideal.[15] Turning to results first from the incidence part of the model, we see that only one variable is statistically significant. As the number of amici filings increases, the likelihood of a precedent being overturned significantly increases. Turning to the duration part of the model, which can be compared to the duration models presented previously, we see some familiar results. First, the conditional impact of ideological distance (when vitality equals 0) is statistically significant. The conditional impact of vitality, when the Court at time *t* is ideological congruent with a precedent, is not statistically significant at the .05 level, but comes close to significance at the .10 level. This evidence, like that from the individual frailty models (in Table 2), refutes Hansford and Spriggs's hypothesis that vitality will significantly lower the risk of an overruling when the Court likes the precedent on ideological grounds. Turning to the interaction term, the results show an insignificant interaction between distance and vitality, again providing evidence against Hansford and Spriggs's hypothesis that increasing vitality significantly amplifies the impact of ideological distance on the risk of an overruling.

Addressing the Issue of Cluster Confounding

Thus far, I have estimated models which do not distinguish the within-precedent and between-precedent effects of ideological distance and vitality, both of which are time-varying covariates. In interpreting the results, neither I (in my previous discussion of results) nor

[15] Perhaps a better specification would be a split population model with a frailty term in the duration part of the model (see Ivanchenko 2006).

Hansford and Spriggs are clear about whether the effects of these two theoretically important variables are between-precedent effects or within-precedent, longitudinal effects. As discussed earlier, when not distinguishing the within versus between cluster variation in a TVC, (1) it is impossible to know the precise nature of the effect, (2) one assumes that the within and between precedent effects are equal, and (3) therefore, one runs the risk of cluster confounding. When operationalizing these TVCs to distinguish within-precedent and between-precedent variation in each, we can make richer, more precise interpretations about their effects. For instance, a significant between-precedent effect of ideological distance would be interpreted as follows: *In comparing precedents, those that are on average more ideologically congruent with the Court will survive significantly longer than those that are ideologically distant*. Note that the between effect is a sort of aggregate comparison of precedents over analysis time. One is interested how the *average* level of the TVC across precedents influences the risk of an overruling. A significant within-precedent, or longitudinal, effect of ideological distance would be interpreted as follows: *Over time, for a given precedent (or within precedents), as ideological distance increases, the risk of an overruling significantly increases*. The within-effect ignores between-precedent variation and analyzes how variation *over time* (within precedents) influences the hazard.

I now address this issue head-on in a model that separates within from between variation in ideological distance and vitality. I examine this issue employing a log-logistic, shared-frailty model, which appeared to provide the best fit of the data compared to the other frailty models. Following the procedures discussed on pp. 13-14, I created both within-precedent and between-precedent operationalizations of ideological distance and vitality. The within and between versions of each are then added to the model. Results from this analysis are presented in Table 5. The first model (the non-interactive model) in Table 5 includes the within and between versions of ideological distance and vitality but does not include any interactions. I return to that issue in the second model in Table 5. The non-interactive results reveal important information about the source of the effects for distance and vitality. Note that the within-precedent effects of both distance and vitality are statistically significant, but the between effects are not statistically significant. This suggests that ideological distance and vitality exhibit longitudinal effects on the hazard of an overruling, but not cross-sectional effects.

[Table 5 about here]

Thus, for ideological distance, we can conclude that over time, for a given precedent, as ideological distance increases, the risk of an overruling significantly increases. The evidence *does not* allow us to conclude the between-precedent effect; that is, in comparing precedents, those that are on average more ideological congruent with the Court *do not* survive significantly longer than those precedents that are more ideologically distant. Results also show that the within and between effects of ideological distance are significantly different (chi-sq=8.94, p<.01). This evidence suggests that the effect of ideological distance reported by Hansford and Spriggs and in the previous models in this paper is the result of cluster confounding; it combines the within and between effects of ideological distance into a "cluster-averaged" effect.

The story is the same for vitality. We can conclude that for a given precedent, as vitality increases over time, the risk of an overruling significantly decreases. But we cannot conclude the between effect—in comparing precedents, those that are on average more vital *do not* survive significantly longer than precedents that are less vital. Moreover, like ideological distance, the within and between effects of vitality are significantly different (chi-sq=4.23, p<.05), suggesting that cluster confounding exists in a model not distinguishing within and between variation in vitality. That is, in the Cox model (from Hansford and Spriggs) and other models presented earlier in this paper, one cannot ascertain precisely whether the effects of vitality are longitudinal or cross-sectional because those models combine, and therefore confound, the two types of effects.

Having now found evidence of distinct within and between effects of the two central variables, what about the issue of the interaction between ideological distance and vitality? To address this issue, I estimated a second model (the interactive model in Table 5) that includes two interaction terms: an interaction between the within-precedent versions of distance and vitality and an interaction between the between-precedent versions of the two variables.[16] Results from Table 5 produce compelling evidence. First, the within interaction is statistically insignificant. Related to Hansford and Spriggs's theory, this suggests that, for a given precedent, increases in vitality *do not* significantly enhance the impact of ideological distance on the risk of an overruling. However, Table 5 reports that the between interaction is statistically significant, adding insight and clarification to Hansford and Spriggs's theory. That is, the result suggests that, in comparing precedents, as average levels of vitality increase across precedents, the impact of the average ideological distance on the risk of an overruling significantly increases.[17] Put another way, for precedents that are more vital on average, the between-effect of ideological distance matters more than it does for precedents that are less vital on average.

CONCLUSION

Duration models with time-varying data can yield important substantive insights about political processes. They can explain how changing circumstances shorten or lengthen the duration of some process. In this paper, I have pointed out what I see as some of the central

[16] One can also imagine that there are cross-level interactions between the two variables, e.g., between the within version of distance and the between version of vitality, etc. I leave this exploration for future research.
[17] Note that the interaction effect is negative, which means that increasing vitality makes the conditional coefficient for ideological distance even more negative (meaning lower survival time and a higher risk of an overruling).

issues and opportunities—of the both the statistical and substantive variety—posed by the inclusion of TVCs. I have suggested that thinking of time-varying duration data as multilevel data—spells nested within subjects—makes one think intently about (1) unobserved heterogeneity and how to account for it via frailty (both unshared and shared) approaches and the split population duration model; and (2) cluster confounding and how to account for it by estimating within-subject, or longitudinal, effects and between-subject effects of TVCs.

The empirical reanalysis of Hansford and Spriggs's examination of the overruling of precedent revealed several insights. Incorporating a frailty model—both unshared and shared—produced evidence that refuted some of Hansford and Spriggs's hypotheses. The models produced evidence of significant heterogeneity, and some of the inferences from these models departed from Hansford and Spriggs's original model that did not account for unobserved heterogeneity. Most prominently, the interaction between ideological distance and vitality was statistically insignificant in all of the frailty models as well as the split population model. Also, addressing the issue of cluster confounding revealed new insights about the precise nature of the effects of ideological distance and vitality. For both variables, results indicated evidence that the within and between-precedent effects were significantly different, thus suggesting that not separating these variables into their within and between components induces cluster confounding. Moreover, the within-precedent effects of both variables were statistically significant, while the between effects were insignificant. In the interactive model, the results again clarified Hansford and Spriggs's theory about the interaction between vitality and distance. The interaction for the between-precedent operationalizations of the variables was statistically significant, while the within interaction was insignificant.

In sum, the issues I have raised about TVCs in duration models can be easily applied by researchers. More importantly, taking advantage of these opportunities brought about by TVCs can further enhance one's understanding of the political process under examination.

REFERENCES

Bafumi, Joseph, and Andrew Gelman. 2006. "Fitting Multilevel Models When Predictors and Group Effects Correlate." Annual Meeting of the Midwest Political Science Association.

Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. "Taking Time Seriously: Time-Series—Cross-Sectional Analysis with a Binary Dependent Variable." *American Political Science Review* 42:1260-88.

Bennett, D. Scott. 1997. "Testing Alternative Models of Alliance Duration, 1816-1984." *American Journal of Political Science* 41:846-78.

Bennett, D. Scott. 1999. "Parametric Models, Duration Dependence, and Time-Varying Data Revisited." *American Journal of Political Science* 43:256-70.

Bennett, D. Scott and Allan C. Stam. 1996. "The Duration of Interstate Wars, 1816-1985." *American Political Science Review* 90:239-57.

Binder, Sarah A., and Forrest Maltzman. 2002. "Senatorial Delay in Confirming Federal Judges, 1947-1998. *American Journal of Political Science* 46:190-99.

Blossfeld, Hans-Peter and Gotz Rohwer. 2002. *Techniques of Event History Modeling*. Mahwah, NJ: Erlbaum.

Box-Steffensmeier, Janet M., Laura W. Arnold and Christopher J. W. Zorn. 1997. "The Strategic Timing of Position Taking in Congress: A Study of the North American Free Trade Agreement." *American Political Science Review* 91(June): 324-338.

Box-Steffensmeier, Janet M., and Bradford S. Jones. 1997. "Time is of the Essence: Event History Models in Political Science." *American Journal of Political Science* 41:1414-61.

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

Box-Steffensmeier, Janet M., and David C. Kimbal. 1999. "The Timing of Voting Decisions in Presidential Campaigns" Annual Meeting of the Midwest Political Science Association.

Box-Steffensmeier, Janet M., Peter M. Radcliffe, and Brandon L. Bartels. 2005. "The Incidence and Timing of PAC Contributions to Incumbent U.S. House Members, 1993-94." *Legislative Studies Quarterly*.

Caldeira, Gregory A., and Christopher Zorn. 2004. "Strategic Timing, Position-Taking, and Impeachment in the House of Representatives." *Political Research Quarterly* 57:517-27.

Cleves, Mario A., William W. Gould, and Roberto G. Gutierrez. 2004. *An Introduction to Survival Analysis Using Stata*. College Station, TX: Stata Press.

Forster, Martin, and Andrew M. Jones. 2001. “The Role of Tobacco Taxes in Starting and Quitting Smoking: Duration Analysis of British Data.” *Journal of the Royal Statistical Society, Series A* 164:517-47.

Hansford, Thomas G., and James F. Spriggs II. 2006. *The Politics of Precedent on the U.S. Supreme Court*. Princeton, NJ: Princeton University Press.

Heckman, James J. 1991. “Identifying the Hand of Past: Distinguishing State Dependence from Heterogeneity.” *American Economic Review* 81:75-79.

Hettinger, Virginia, and Christopher Zorn. 2005. “Explaining the Incidence and Timing of Congressional Responses to the U.S. Supreme Court.” *Legislative Studies Quarterly* 30:5-28.

Hougaard, Philip. 2000. *Analysis of Multivariate Survival Data*. New York: Springer-Verlag.

Ivanchenko, Roman. 2006. “Congress and Supreme Court Over Time: Using a Frailty Model in a Split Population Analysis.” Presented at the Annual Meeting of the American Political Science Association.

King, Gary, James Alt, Nancy Burns, and Michael Laver. 1990. “A Unified Model of Cabinet Dissolution in Parliamentary Democracies.” *American Journal of Political Science* 34:847-71.

Kuk, A.Y.C., and C.H. Chen. 1992. “A Mixture Model Combining Logistic Regression with Proportional Hazards Regression.” *Biometrika* 79:531-41.

Scherer, Nancy, Brandon L. Bartels, and Amy Steigerwalt. 2007. “Sounding the Fire Alarm: The Role of Interest Groups in the Lower Court Confirmation Process.” Annual Meeting of the Southern Political Science Association.

Schmidt, Peter, and Ann D. Witte. 1984. *An Economic Analysis of Crime and Justice: Theory, Methods, and Applications*. Orlando, FL: Academic Press.

Schmidt, Peter, and Ann D. Witte. 1989. “Predicting Criminal Recidivism Using ‘Split Population’ Survival Time Models.” *Journal of Econometrics* 40:141-59.

Segal, Jeffrey A., and Harold J. Spaeth. 2002. *The Supreme Court and the Attitudinal Model Revisited*. New York: Cambridge University Press.

Skrondal, Anders, and Sophia Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman & Hall.

Spriggs, James F., and Thomas G. Hansford. 2001. “Explaining the Overruling of U.S. Supreme Court Precedent.” *Journal of Politics* 63:1091-1111.

Sy, Judy P., and Jeremy G. Taylor. 2000. “Estimation in a Cox Proportional Hazards Cure Model.” *Biometrics* 56:227-36.

Vaupel, J.W., Manton, K.G., Stallard, E. 1979. “The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality.” *Demography* 16:439-54.

Zorn, Christopher. 2000. “Modeling Duration Dependence.” *Political Analysis* 8:367-80.

Zorn, Christopher. 2001. “Estimating Between- and Within-Cluster Covariate Effects, with an Application to Models of International Disputes.” *International Interactions* 27:433-45.

Table 1: Example of Duration Data with Time-Varying and Time-Constant Covariates

Obs. ID	Subject	Spell	Event	X1	X2	X3
1	1	1	0	10	0	25
2	1	2	0	8	0	25
3	1	3	0	2	0	25
4	1	4	0	3	0	25
5	2	1	0	1	1	19
6	2	2	0	5	1	19
7	2	3	0	6	1	19
8	2	4	0	7	1	19
9	2	5	0	6	1	19
10	2	6	1	9	1	19
11	3	1	0	0	1	10
12	3	2	0	4	1	10
13	3	3	0	1	1	10
14	3	4	1	2	1	10
15	4	1	0	10	0	14
16	4	2	0	5	0	14
17	5	1	0	9	1	32
18	5	2	0	8	1	32
19	5	3	0	3	1	32
20	5	4	0	5	1	32

Table 2: Individual (Spell-Specific) Frailty Models of the Overruling of Precedent

	Hansford and Spriggs's Cox Model (Non-frailty)			Weibull Model with Individual (Spell-specific) Frailty			Log-Logistic Model with Individual (Spell-specific) Frailty		
	Coeff.	**(SE)**	**p**	**Coeff.**	**(SE)**	**p**	**Coeff.**	**SE**	**p**
Ideological Distance	0.039	(0.009)	0.000	-0.045	(0.014)	0.001	-0.043	(0.013)	0.001
Vitality	-0.390	(0.151)	0.005	0.160	(0.205)	0.218	0.179	(0.192)	0.175
Ideological Distance* Vitality	0.009	(0.005)	0.025	0.008	(0.013)	0.269	0.005	(0.011)	0.321
Control Variables									
Concurring Opinions	0.228	(0.189)	0.115	-0.345	(0.204)	0.046	-0.327	(0.195)	0.047
Voting Margin	-0.164	(0.036)	0.000	0.199	(0.051)	0.000	0.200	(0.050)	0.000
Total Prior Interpretations	0.084	(0.028)	0.001	-0.306	(0.101)	0.001	-0.280	(0.093)	0.001
Court Agenda	0.016	(0.009)	0.041	-0.025	(0.013)	0.023	-0.025	(0.012)	0.019
Breadth	0.397	(0.151)	0.004	-0.358	(0.237)	0.066	-0.340	(0.223)	0.064
Amici Filings	0.117	(0.064)	0.034	-0.070	(0.120)	0.280	-0.083	(0.114)	0.233
Media Coverage	-0.069	(0.254)	0.393	-0.006	(0.333)	0.493	-0.002	(0.321)	0.498
Per Curiam Opinion	-0.279	(0.526)	0.298	0.368	(0.586)	0.265	0.312	(0.572)	0.293
Constitutional Case	0.557	(0.221)	0.006	-0.714	(0.269)	0.004	-0.698	(0.261)	0.004
Constant	-	-	-	8.015	(0.789)	0.000	8.130	(0.778)	0.000
	LL=-836.75 Chi-sq=142.02, p<.001			LL=-554.72 Chi-sq=133.37, p<.001 *p* (shape param.)=1.086 θ=13.432 (Chi-sq=21.33, p<.001) AIC=1139.433			LL=-554.92 Chi-sq=132.71, p<.001 γ=0.951 θ =8.304 (Chi-sq=14.16, p<.001) AIC=1139.847		

For all models: number of spells=179,653; number of precedents=6,363; p-values are one-tailed, in accord with Hansford and Spriggs

Table 3: Shared (Precedent-Specific) Frailty Models of the Overruling of Precedent

	Hansford and Spriggs's Cox Model (Non-frailty)			Weibull Model with Shared (Precedent-specific) Frailty			Log-Logistic Model with Shared (Precedent-specific) Frailty		
	Coeff.	**(SE)**	**p**	**Coeff.**	**(SE)**	**p**	**Coeff.**	**SE**	**p**
Ideological Distance	0.039	(0.009)	0.000	-0.035	(0.010)	0.001	-0.037	(0.012)	0.001
Vitality	-0.390	(0.151)	0.005	0.316	(0.159)	0.024	0.266	(0.171)	0.060
Ideological Distance* Vitality	0.009	(0.005)	0.025	-0.001	(0.007)	0.428	0.002	(0.009)	0.403
Control Variables									
Concurring Opinions	0.228	(0.189)	0.115	-0.338	(0.202)	0.047	-0.341	(0.200)	0.045
Voting Margin	-0.164	(0.036)	0.000	0.209	(0.048)	0.000	0.210	(0.050)	0.000
Total Prior Interpretations	0.084	(0.028)	0.001	-0.230	(0.063)	0.000	-0.254	(0.079)	0.001
Court Agenda	0.016	(0.009)	0.041	-0.027	(0.010)	0.004	-0.028	(0.011)	0.005
Breadth	0.397	(0.151)	0.004	-0.393	(0.232)	0.045	-0.360	(0.223)	0.054
Amici Filings	0.117	(0.064)	0.034	-0.078	(0.120)	0.259	-0.071	(0.119)	0.275
Media Coverage	-0.069	(0.254)	0.393	0.030	(0.326)	0.463	0.006	(0.324)	0.493
Per Curiam Opinion	-0.279	(0.526)	0.298	0.382	(0.560)	0.248	0.352	(0.557)	0.264
Constitutional Case	0.557	(0.221)	0.006	-0.693	(0.259)	0.004	-0.698	(0.261)	0.004
Constant	-	-	-	7.339	(0.719)	0.000	7.566	(0.728)	0.000
	LL=-836.75 Chi-sq=142.02, p<.001			LL=-551.68 Chi-sq=139.44, p<.001 *p* (shape param.)=1.198 θ=28.497 (Chi-sq=27.39, p<.001) AIC=1133.367			LL=-550.94 Chi-sq=140.68, p<.001 gamma=0.863 θ=21.918 (Chi-sq=22.13, p<.001) AIC=1131.879		

For all models: number of spells=179,653; number of precedents=6,363; p-values are one-tailed, in accord with Hansford and Spriggs

Figure 1: Survival Curves for High and Low Values of Ideological Distance and Vitality; from Log-Logistic, Shared Frailty Model

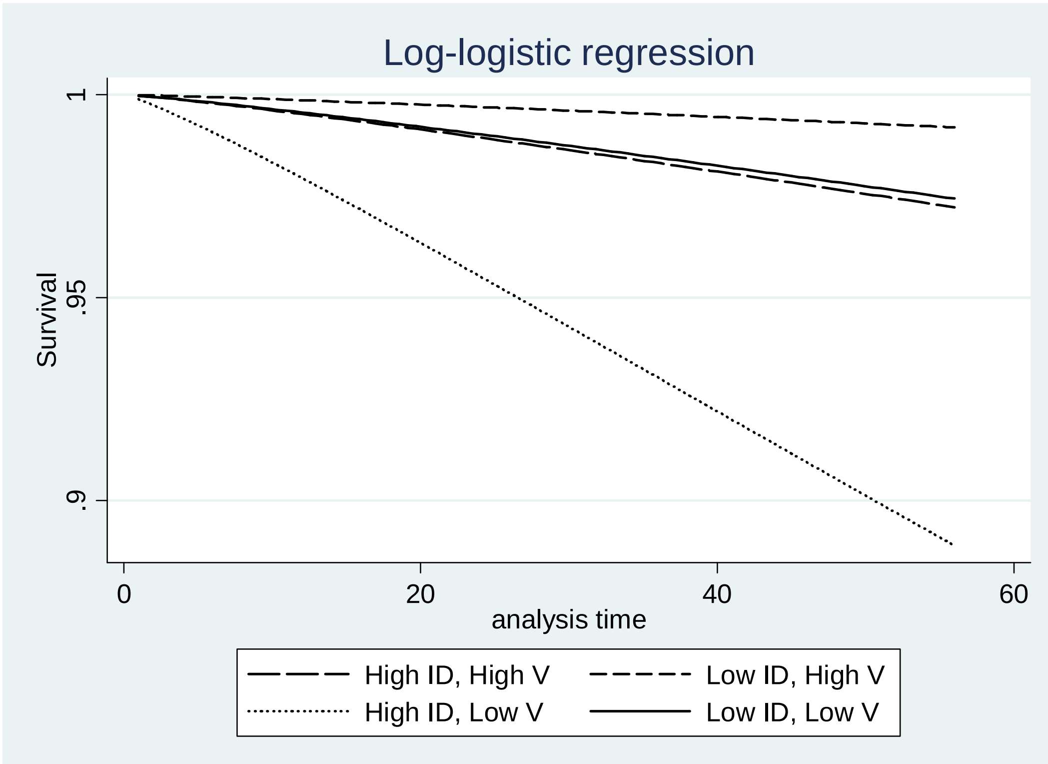

Table 4: Split Population Duration Model of the Overruling of Precedent

	Likelihood of Overruling			Timing of Overruling		
	Coeff.	**(SE)**	**p**	**Coeff.**	**(SE)**	**p**
Ideological Distance	-	-	-	-0.042	(0.014)	0.001
Vitality	-	-	-	0.227	(0.193)	0.120
Ideological Distance* Vitality	-	-	-	0.007	(0.012)	0.287
Control Variables						
Concurring Opinions	-0.040	(0.134)	0.767	-0.433	(0.297)	0.072
Voting Margin	-0.063	(0.053)	0.232	0.087	(0.110)	0.215
Total Prior Interpretations	-	-	-	-0.309	(0.088)	0.000
Court Agenda	-	-	-	-0.023	(0.013)	0.031
Breadth	-0.034	(0.148)	0.818	-0.522	(0.339)	0.062
Amici Filings	0.325	(0.151)	0.032	0.408	(0.177)	0.011
Media Coverage	-0.004	(0.242)	0.987	0.144	(0.533)	0.394
Per Curiam Opinion	0.662	(0.941)	0.482	1.221	(1.127)	0.140
Constitutional Case	-0.425	(0.441)	0.336	-1.565	(0.771)	0.021
Constant	-0.600	(0.538)	0.265	6.894	(1.134)	0.000

LL=-820.87; Chi-sq=30.74, p<.01; γ=.91

Est. Split: 0.169; Actual Split=0.017

Number of spells=179,653; number of precedents=6,363; p-values are one-tailed in the timing part, and two-tailed in the likelihood part.

Table 5: Log-Logistic Shared Frailty Models Assessing the Within and Between Effects of Ideological Distance and Vitality

	Non-interactive model			Interactive model		
	Coeff.	**(SE)**	**p**	**Coeff.**	**(SE)**	**p**
Ideological Distance (Within)	-0.072	(0.018)	0.000	-0.073	(0.019)	0.000
Ideological Distance (Between)	-0.001	(0.015)	0.484	-0.004	(0.015)	0.398
Vitality (Within)	0.567	(0.172)	0.001	0.585	(0.191)	0.001
Vitality (Between)	0.096	(0.144)	0.252	0.540	(0.274)	0.025
Ideo. Dist.*Vitality (Within)				0.011	(0.024)	0.329
Ideo. Dist.*Vitality (Between)				-0.037	(0.019)	0.026
Control Variables						
Concurring Opinions	-0.280	(0.188)	0.068	-0.284	(0.187)	0.064
Voting Margin	0.226	(0.049)	0.000	0.229	(0.050)	0.000
Total Prior Interpretations	-0.224	(0.073)	0.001	-0.226	(0.072)	0.001
Court Agenda	-0.034	(0.011)	0.001	-0.034	(0.011)	0.001
Breadth	-0.381	(0.216)	0.039	-0.366	(0.216)	0.045
Amici Filings	-0.097	(0.111)	0.192	-0.104	(0.110)	0.173
Media Coverage	-0.048	(0.308)	0.439	-0.048	(0.308)	0.438
Per Curiam Opinion	0.247	(0.554)	0.328	0.232	(0.561)	0.340
Constitutional Case	-0.754	(0.253)	0.002	-0.793	(0.259)	0.001
Constant	7.577	(0.697)	0.000	7.725	(0.712)	0.000
	LL=-544.25 Chi-sq=154.05, p<.001 γ=.92 θ=13.36 (Chi-sq=10.42, p<.001)			LL=-542.05 Chi-sq=158.45, p<.001 γ=.95 θ=11.60 (Chi-sq=8.22, p<.01)		

Number of spells=179,653; number of precedents=6,363; p-values are one-tailed

Using data from the Supreme courts of the United States, Massachusetts, and Canada we show that each court's docket features a slow decay with a decreasing tail. This demonstrates that, in each of the courts examined, the vast majority of cases are resolved relatively quickly, while there remains a small number of outlier cases that take an extremely long time to resolve. We discuss the implications for this on legal systems, the study of the law, and future research.Â We will then describe our data collection and analysis methods followed by a model for court case priority queueing, an empirical demonstration, and finally a brief discussion of the implications these findings have. Overview. 2 Network Processes in Supreme Court Citations. 3 The Citation Exponential Random Graph Model. 4 Empirical Analysis. 5 Conclusion. A cERGM Estimation.Â Abstract The significance and influence of US Supreme Court majority opinions derive in large part from opinionsâ€™ roles as precedents for future opinions. A growing body of literature seeks to understand what drives the use of opinions as precedents through the study of Supreme Court case citation patterns. We raise two limitations of existing work on Supreme Court citations. First, dyadic citations are typically aggregated to the case level before they are analyzed. Time-varying covariates make this a little bit more complicated. To use a time-varying covariate, you must divide a customer's lifetime into "chunks" where the various values of the covariates apply. For example, check out this snippet of data below that includes survival data, plus an indicator showing whether a customer has contacted support: Instead of simply an end time and a churn indicator, we now have an additional start time variable.Â For now, on to the modeling! If you'd like to work with the full set of dummy data used for this post, you can grab it here. Doing some analysis! Once you have your data set up, doing the actual cox regression looks pretty much like doing any other cox regression. Supreme Court clerkship, law clerks can virtually dictate their own career path. For instance, clerks can work at the Department of JusÂ 40 At the same time, those Justices vary widely in the degree of additional responsibility they delegate to their law clerks; moreover, minor variations exist across chambers in the details of how cert and bench memoranda are prepared. 41. The extent and nature of Supreme Court clerks' formal responsibili-ties is directly relevant to the question of their possible influence.