Modeling Heterogeneity in Duration Models

Janet M. Box-Steffensmeier
Department of Political Science
Ohio State University
Columbus, OH 43210-1373
jboxstef+@osu.edu

Christopher Zorn
Department of Political Science
Emory University
Atlanta, GA 30322
czorn@emory.edu

Version 1.0
July 12, 1999

Paper prepared for presentation at the 1999 Summer Meeting of the Political Methodology Society, July 15-17, 1999, Texas A&M University, College Station, TX. Thanks to Mario Cleves, Dan Reiter, Narayan Sastry, and Terry Therneau for helpful discussions, and Kevin Sweeney and Andy Tomlinson for research assistance. This paper is available at the Political Methodology website (http://polmeth.calpoly.edu).

Abstract

As increasing numbers of political scientists have turned to event history models to analyze duration data, there has been growing awareness of the issue of heterogeneity: instances in which subpopulations in the data vary in ways not captured by the systematic components of standard duration models. We discuss the general issue of heterogeneity, and offer techniques for dealing with it under various conditions. One special case of heterogeneity arises when the population under study consists of one or more subpopulations which will never experience the event of interest. Split-population, or "cure" models, account for this heterogeneity by permitting separate analysis of the determinants of whether an event will occur and the timing of that event, using mixture distributions. We use the split-population model to reveal additional insights into the strategies of political action committees' allocation decisions, and compare split-population and standard duration models of Congressional responses to Supreme Court decisions.

We then go on to explore the general issue of heterogeneity in survival data by considering two broad classes of models for dealing with the lack of independence among failure times: variance correction models and "frailty" (or random effects) duration models. The former address heterogeneity by adjusting the variance matrix of the estimates to allow for correct inference in the presence of that heterogeneity, while the latter approach treats heterogeneity as an unobservable, random, multiplicative factor acting on the baseline hazard function. Both types of models allow us to deal with heterogeneity that results, for example, from correlation at multiple levels of data, or from repeated events within units of analysis. We illustrate these models using data on international conflicts.

In sum, we explore the issue of heterogeneity in event history models from a variety of perspectives, using a host of examples from contemporary political science. Our techniques and findings will therefore be of substantial interest to both political methodologists and others engaged in empirical work across a range of subfields.

Not chaos-like together crush'd and bruis'd,
But as the world, harmoniously confus'd,
Where order in variety we see,
And where, though all things differ, all agree.

- Alexander Pope

1. Introduction

Statisticians and researchers have long been aware of the ill effects of unobserved heterogeneity in regression models. In the literature on linear models, these effects are well-known (e.g. Judge et. al. 1985, Chapter 13): disregarding group- or individual-specific variation is a special case of the omitted variables problem, with its resulting biases. Extensions to other models are also widespread; Gourieroux, Monfort, and Trognon (1984), for example, show that if the assumption of homogeneity is incorrect, the parameter estimates of a Poisson model will be inconsistent and/or inferences will be based on inappropriate standard errors. Likewise, in the case of heterogeneity in duration models, "if the model specification is incomplete and if systematic individual differences in the distribution remain after the observed effects are accounted for, then inference based on the improperly specified model is likely to be problematic" (Greene 1997, 995; see also Omori and Johnson 1993).

But while its presence can be problematic in any model, unobserved heterogeneity is of particular concern in models of duration data. A simple illustration of why this is the case is to consider a population consisting of two subpopulations with different risks of experiencing the event in question. Figure 1 illustrates two groups with different, time-constant (i.e., exponential) hazards, along with the estimated hazard in a model which fails to account for the difference between the groups. Because the more failure prone are removed at a higher rate than the less failure-prone, the proportion of the former to the latter in the sample declines over time. As a result, the failure rate in the surviving population will appear to fall over time, despite the fact that the hazards for both groups remain constant over time. In other words, unobserved heterogeneity may give the appearance of an aggregate decline in hazard rates, simply because the high-risk observations have already experienced the event and thus exited, despite the fact that individual hazards are constant or even rising.[1]

One obvious lesson of this example is that we should be concerned with model specification: proper inclusion of relevant covariates is the single most important step one can take to reduce the

[1]See Proschan (1963) for a formal proof that a mixture of two exponential distributions with different failure rates does not result in an exponential distribution.

deleterious effects of heterogeneity. This control may be incomplete, however, if, for example, covariates are inappropriately omitted, the functional form is misspecified, unobservable variables are important, or variable effects vary across members of the sample (Pickles and Crouchley 1995). As suggested above, the most common problem resulting from ignored heterogeneity is that the estimated hazard rate becomes biased toward negative duration dependence (Heckman and Singer 1984).[2] Aalen's (1992) important article points out that the distortion of survival curves and the hazard rate has been discussed by a number of authors, including Manton, Stallard, and Vaupel (1981), Vaupel and Yashin (1985), Hougaard (1984, 1986a, 1986b), Aalen (1988) and Vaupel (1990).

Here we consider the issue of heterogeneity in survival models, and offer overviews and examples of techniques with different forms of that heterogeneity. Generally speaking, we will consider hazard models with heterogeneity in the form of individual-, cluster- or group-level effects, α_i, not accounted for by the independent variables:

$$h(t) = f\,(\mathbf{X}_i\beta,\alpha_i) \qquad \textbf{(1.1)}$$

This general formulation encompasses a range of interesting cases. We begin with a simple case of heterogeneity arising when the population under study consists of one or more subpopulations which will never experience the event of interest. Split-population models account for this heterogeneity by permitting separate analysis of the determinants of whether an event will occur and the timing of that event, using mixture distributions. We go on to explore the general issue of heterogeneity in survival data by considering two broad types of approaches for dealing with the lack of independence among failure times: variance correction models and "frailty" (or random effects) duration models.

[2] In the linear model, if the excluded independent variables follow a normal distribution, the distribution of the dependent variable is still normal but the variance is increased. The problem with event history models is that the excluded independent variables lead to distributions outside the family considered (Hougaard 1999, 21). "For example, considering neglected covariates in a constant hazard model leads to a distribution with a decreasing hazard . . . a major drawback of the proportional hazards model is that the model and the value of the relative risk are not robust toward neglected covariates" (1999, 21-2).

2. Split-Population Survival Models

A simple form of heterogeneity along the lines discussed above occurs when one subpopulation has a hazard which is zero; i.e., they will never experience the event in question. Such heterogeneity is relatively common in social scientific studies, as well as in other areas where survival models are widely used (e.g. biometric and medical studies). Schmidt and Witte (1988), for example, discuss the example of criminal recidivism, where researchers are often interested in modeling the effect of in-prison programs on the behavior of convicts upon their release. Implicit in standard survival models is the assumption that all observations will "fail"; i.e., that every individual released from prison will eventually return to prison. Similarly, in dyadic studies of international conflict, an unstated assumption of conventional approaches is that, given an adequate follow-up time, all dyads would eventually go to war.

Split-population models are also known as "cure models" in the biometrics literature, i.e., part of the population is cured and will never experience the event, and have both a long history (e.g. Boag 1949; Berkson and Gage 1952) and widespread applications and extensions in recent years (e.g. Maltz and McCleary 1977; Vaupel et. al. 1979; Farewell 1982; Aalen 1988, 1992; Kuk and Chen 1992; Longini and Halloran 1996; Tsodikov 1998; Tsodikov et. al. 1998). The intuition behind these models is that, while standard duration models require a proper distribution for the density which makes up the hazard (i.e., one which integrates to one; in other words, that all subjects in the study will eventually fail), split population models allow for a subpopulation which never experiences the event of interest. This is typically accomplished through a mixture of a standard hazard density and a point mass at zero (Maller and Zhao 1996). That is, split population models estimate an additional parameter (or parameters) for the probability of eventual failure, which can be less than one for some portion of the data. In contrast, standard event history models assume that eventually *all* observations will fail, a strong and often unrealistic assumption.

We present the mathematical details of a general parametric split population model and illustrate the usefulness of these models with two applications, both involving processes in which we would not expect all observations to experience the event of interest. The first topic is campaign finance; specifically, the timing of campaign contributions from political action committees (PACs) to members of Congress. The second example considers congressional responses to Supreme Court decisions.

2.1 A General Split-Population Model for Survival Data[3]

We begin our discussion with a standard parametric model for continuous-time duration data, where the duration of interest t is assumed to have a distribution function $f(t,\theta)$, with θ a parameter vector to be estimated. Define $F(t,\theta) = \Pr(T \leq t)$, $t > 0$ as the corresponding cumulative density, where T represents the duration defined by the end of the observation period. The associated survival function (defined as the probability of survival to time t and denoted $S(t,\theta)$) is then equal to $1-F(t,\theta)$. If $f(t,\theta)$ is the density function of $F(t,\theta)$, then we can write the hazard rate:

$$h(t,\theta) = f(t,\theta) / S(t,\theta) \qquad \textbf{(2.1)}$$

This value is akin to the conditional probability of a failure at time t given that no failure has occurred prior to t (e.g. Box-Steffensmeier and Jones 1997).

We consider a model for the duration t which splits the sample into two groups, one of which will eventually experience the event of interest (i.e., "fail") and the other which will not. We define the latent variable Y such that $Y_i = 1$ for those who will eventually fail and $Y_i = 0$ for those who will not; define $\Pr(Y_i = 1) = \delta_i$. The conditional density and distribution functions are defined as:

$$f\,(t_i * Y_i = 1) = g(t,\theta)$$
$$F\,(t_i * Y_i = 1) = G(t,\theta),$$

while leaving $f\,(t * Y_i = 0)$ and $F\,(t * Y_i = 0)$ undefined.[4] Let R_i be the observable indicator of failure, such that $R_i = 1$ when failure is observed and $R_i = 0$ otherwise. For those observations which fail during the observation period, we observe $R_i = 1$ and their duration. Since these observations also necessarily have $Y_i = 1$, we can write the unconditional density for these observations as:

$$\Pr(Y_i=1)\Pr(t_i \leq T_i * Y_i=1) = \delta_i\, g(t_i,\theta) \qquad \textbf{(2.2)}$$

[3] This section draws extensively on Schmidt and Witte (1989).

[4] Because $Y_i = 0$ implies that the observation will never fail (and thus the duration will never be observed), the probabilities for $g(t_i * Y_i=0)$ and $G(t_i * Y_i=0)$ cannot be defined.

where T_i indicates the censoring time. In contrast, for those observations in which we do not observe a failure ($R_i = 0$), this may be due to either $Y_i = 0$ (the observation will never fail) or because $t_i > T_i$. (the observation is censored). The unconditional density for observations with $R_i = 0$ is therefore:

$$\Pr(Y_i = 0) + \Pr(Y_i = 1)\Pr(t_i > T_i * Y_i = 1) \quad = \quad (1 - \delta_i) + \delta_i G(T_i, \theta) \tag{2.3}$$

Combining these values for each of the respective sets of observations, and assuming independence across observations, the likelihood function is:

$$L = \prod_{i=1}^{N} \delta_i g(t_i, \theta)^{R_i} [1 - \delta_i + \delta_i G(t_i, \theta)]^{(1-R_i)} \tag{2.4}$$

with the corresponding log-likelihood:

$$\ln L = \sum_{i=1}^{N} R_i [\ln \delta_i + \ln g(t_i, \theta)] + (1 - R_i) \ln[1 - \delta_i + \delta_i G(T_i, \theta)] \tag{2.5}$$

The probability δ_i is typically modeled as a logit, although other specifications (e.g. probit, complimentary log-log, etc.) are also possible, and can include explanatory variables:[5]

$$\delta_i = \frac{exp(Z_i\gamma)}{1 + exp(Z_i\gamma)} \tag{2.6}$$

Note that when $\delta = 1$ (i.e. when all observations will eventually fail), the equation reduces to the standard general duration model with censoring.[6] The application of split population models should be considered whenever all observations cannot reasonably be assumed to fail at some point in the future. A useful property of the split-population model is that it allows for separate estimation of the influence of covariates on the probability of being in an "immune" class from their effect on the time until the event of interest for non-immunes. In this respect, these models are quite similar to "mover-stayer" models (e.g. Yamaguchi

[5] Note that this model is identified even when the variables in δ_i are identical to those in the model of duration. This means that one can test for the effects of the same set of variables on both the incidence of failure and the duration associated with it (Schmidt and Witte 1989).

[6] Note, however, that testing $\delta = 1$ is generally not possible, as it represents a border case; moreover, it does not correspond to the case of $Z_i\gamma = 0$ (which yields $\delta = 0.5$).

1992) and to “hurdle” and “zero-inflated” event count models (e.g. Zorn 1998).

2.2 Example One: Congress, PACs, and the Strategy of Campaign Finance

As has been demonstrated in recent years (e.g. Box-Steffensmeier 1996), duration models offer the potential for important insights into the issue of campaign finance. This may be particularly true in the area of campaign contributions, where (as recent events attest) timing can be of critical importance. Yet conventional duration models of campaign contributions would assume that, eventually, every PAC or other group will contribute to every member of Congress, clearly an implausible assumption to make. Box-Steffensmeier and Radcliffe (1996) analyze data on the timing of contributions where there is a record for every possible combination of a PAC and a candidate pair to examine both whether a contribution was given and if so, when. Covariate data on a number of factors previous studies have found to be influential on PAC contributions (e.g. voting patterns and electoral vulnerability) are also included in the models. Here, some estimation issues and interpretation of the split population model are illustrated with this example.

The most commonly used parametric survival time distributions are the normal, logistic, exponential, and Weibull.[7] In the context of the timing of PAC contributions, there is little theoretical reason for preferring one distribution over another; accordingly, we follow Blossfeld and Rohwer (1995, 212) and Schmidt and Witte (1988, 98) and opt for the log-logistic model on the basis of fit.[8] In a split population model, one would prefer the estimated splitting parameter, δ (which represents the probability of eventually giving a contribution), to be as close as possible to the observed failure rate. This estimated failure rate or split and the observed split are at the bottom of Table 1. The models do very well in estimating the actual contribution “failure” rate, for example, it is estimated as .26 and observed as .24 for

---

[7] For a discussion of a nonparametric split population model see Price (1999) and of a split population model with time varying covariates, see Tsodikov, Loeffler, and Yakovlev (1998) .

[8] The models for small and large corporate PACs are log-logistic. The hazard for the log-logistic model is defined as $h(t) = \dfrac{\lambda_i \alpha (\lambda_i t)^{\alpha-1}}{1+(\lambda_i t)^{\alpha}}$, and “is even more flexible than the Gompertz and Weibull distributions” (Blossfeld and Rohwer 1995, 183) in that it may yield either monotonic or nonmonotonic hazards, depending on the shape parameter α. Other distributions considered included the exponential, Weibull (which is an encompassing model for the logistic), normal, and logistic. Kuk and Chen (1992) introduced a split-population variant of the Cox proportional hazards model, but it has received relatively little use in comparison to the parametric models.

large corporate PACs.

Table 1 contains the timing and likelihood results for large and small corporate PACs. For the timing model, positive (negative) coefficients lead to later (earlier) contributions. For the probability model, we estimate the likelihood that a PAC-candidate pair will never fail, i.e., exchange a contribution, so negative (positive) coefficients produce an increase (reduction) in the likelihood of a contribution being forthcoming. Together with the significance values, the results give a picture of the determinants of both contributions and timing. Recall that split population models allow one to test whether the event, the giving of a contribution, and the timing of the event depend on different factors. That is, separate tests may be undertaken to determine whether the factors that explain receiving a contribution are different from the factors that explain the timing of receiving a contribution.

More covariates help explain corporate PAC contribution strategies than the timing of those contributions. For example, a seat on the Energy and Commerce Committee positively affects the probability of a contribution, but not its timing. A PACs' resources and geography are overwhelmingly central to both the likelihood and timing of a contribution while candidate power, ideology, and need all play less of a role in timing than in contributing. The factors affecting the likelihood of a contribution are very similar for large and small corporate PACs; there is more variation between the timing models for large and small corporate PACs.

A seat on the Energy and Commerce committee served as a powerful lure for contributions, from both large and small PACs. A prestige committee assignment likewise increased the likelihood of a contribution from both large and small corporate PACs. However, neither of these positions had a statistically significant impact at the five percent level on the timing of those contributions. Senior members likewise had a statistically significant greater chance of receiving a contribution. However, more senior members were given their contributions later than their more junior counterparts. Members of the party leadership were also more likely to receive contributions, but only Democratic majority party leaders received those contributions earlier in the cycle. Positions of power increased the likelihood of contributions, but not necessarily early ones.

Democratic backbenchers profited from their majority status in 1993-94 along with their party leaders. Democrats were more likely to receive contributions and receive them early, from both large and small corporate PACs. However, Republicans could overcome the party effect with a business-friendly record. Higher COC scores increased the likelihood of contributions from both large and small corporate PACs. From the relative coefficient magnitudes, both large and small corporate PACs placed greater emphasis on ideology, as the coefficient for the 0-1 scaled COC rating scale is far larger than that of the dichotomous party variable for the likelihood of receiving a contribution. Incumbents who were elected by narrower margins in the previous election or faced a quality challenger in the primary or general election were more likely to receive contributions from both large and small corporate PACs. There is some evidence that corporate PACs reduced risk by staying out of competitive races as shown by later contributions when there was a quality challenger in the general election or if the member was in a close race in the previous election.

Financial need factors cut in different directions. Members of Congress representing wealthier districts were no more likely to receive contributions from large or small corporate PACs. But those who relied more heavily on PAC contributions were more likely to receive contributions and to receive them earlier from both large and small corporate PACs. Beginning cash-on-hand has no statistically significant effect for small PACs, while large PACs were more likely to contribute, but to do so later in the campaign. Even within the size categories, size matters. Larger PACs in each category were more likely to make contributions, and to make them earlier. Every decision a PAC makes is contingent on their supply of funds. There is evidence of declining marginal returns for both small and large PACs; the square of receipts is statistically significant and the opposite sign of receipts. Finally, a powerful role is played by the geography of the PAC's donors. The larger the share of a PACs fundraising in a state, the more likely they were to make contributions in that state, and to make them earlier.

Judging the relative importance of covariates in non-linear models requires some care, particularly in models with both dichotomous and continuous covariates. The most straightforward method is to use fitted values (see King 1989). We compare a baseline case to four alternatives in Table 2.[9] The first alternative case, "power" represents an otherwise typical member who has twenty years of seniority (one standard deviation above the mean) and is part of the Democratic party leadership. In the second alternative case, "ideology" represents an otherwise typical member whose COC or COPE score for corporate and labor PACs, respectively, is seventy-five rather than the baseline of fifty. "Need" represents an otherwise typical member who represents a marginal district (elected with one standard deviation below the mean for previous vote), and who relies more heavily on PACs as a source of campaign funds (one standard deviation above the mean). Finally, "geography" represents an otherwise typical member whose

[9] The baseline case was created with all the dichotomous variables set to zero and all the continuous variables set to their mean, except the ideology scores, which were set to the middle of the range at 50.

state provided ten percent (one standard deviation above the mean) of the relevant PAC's traceable contributions, and for labor PACs contributions also represented a district containing a greater concentration of labor union members (one standard deviation above its mean) among its residents.

Table 2 provides estimates of the probability of receiving a contribution for each of the cases from each type of PAC. The difference between large and small PACs is immediately apparent. While the probability of receiving a contribution from a large corporate PAC varies between .18 and .30, for small PACs that probability lies between .03 and .04. This emphasizes the problem of pooling all corporate PACs, for example. For large corporate PACs, party leaders fare the best, with just under a .30 chance of receiving a contribution. Geography matters least, improving only .03 on the baseline probability. For small corporate PACs, the pattern is similar. The party leader has just under a .04 chance of receiving a contribution. The most needy candidate finishes last with just over a .03 chance.

The effects of timing for each of the example cases also can be examined by considering the estimated baseline hazard rate. Figure 2 plots the baseline hazards (i.e., the probability of a PAC and candidate pair exchanging a contribution on a particular day, given that they have not yet done so) for small and large corporate PACs. The hazard rate for the large PACs peak and then decline late in the first year, while the small PACs peak later in the second year and exhibit a correspondingly smaller decline.

2.3 Example Two: Congressional Responses to Supreme Court Decisions, 1979-88

Our second example illustrates the potential split-population models offer for providing greater insight into political processes than do conventional duration models. We take as our example here one aspect of the separation of powers: specifically, the issue of Congressional responses to decisions of the U.S. Supreme Court, in the form of bills, hearings, or other kinds of formal actions taken in response to Court decisions. Such responses have two signature characteristics: they are typically taken in order to modify or reverse the Court's decisions, and they are relatively rare. In our data,[10] for example, only 132 of the 7033 decisions under scrutiny (1.9 percent) were the target of Congressional responses. But while it is unlikely that most Court decisions will ever be subject to Congressional scrutiny, scholars remain

[10] Specifically, we examine Congressional responses to the decisions of the Warren and Burger Courts, i.e., the 1953 to 1985 terms, taken during the 96th-100th Congresses (1979-1988), as reported in Eskridge (1991). There are 7157 such cases; omitting 124 cases because of missing data (mostly on the "Liberal Decision" variable) yields a total of 7033 cases for analysis. For a more thorough analysis of such responses, see Zorn and Caldeira 1995; for a similar split-population analysis of successful Congressional *overrides* of Supreme Court statutory decisions, see Hettinger and Zorn 1999.

interested in those cases which will, and in the conditions under which those responses occur. Split-population models are ideal for this kind of analysis, in that they allow us to separate the effects of case-specific independent variables on the probability of the case *ever* being subject to Congressional scrutiny from the timing of that response.

Our dependent variable is thus the duration in years between the decision of the Court and the first Congressional response.[11] As noted above, only 2 percent of the cases in our data experience such responses; the remainder are censored. We examine the influence of a number of independent variables on the hazard of a response, including the year of the decision itself; the presence (coded 1) or absence (coded 0) of disagreement in the lower court decision, alteration of precedent by the Supreme Court, a declaration of unconstitutionality, or a liberal policy decision by the Court; the number of briefs *amicus curiae* filed on the merits in the case; and a series of indicator variables for the nature of the losing party to the case (federal, state, and local governments, businesses, class action litigants, and natural persons, with non-profit groups omitted as the baseline category).[12] In general, these variables indicate the salience of the case to members of Congress, either through their inherent importance in the constitutional system or their impact on important constituent groups or actors in the Congressional arena. We estimate two models: a standard log-logistic hazard model, and a split-population model which is log-logistic in the duration and uses a probit link for the probability of no response. Results of these estimates are presented in Table 3.

The standard log-logistic model yields results indicating that only two variables (year of decision and *amicus* briefs) significantly affect the hazard of a response, though several others (lower court disagreement, declaration of unconstitutionality, and state and natural person losing parties) are of marginal significance ($p < .10$, two-tailed). The estimated s parameter, which (as in the Weibull model) indicates the extent of duration dependence, is not significantly different from 1.0 ($z = 0.843$), indicating that the hazard of a Congressional response, conditional on the independent variables and coefficient estimates, remains relatively constant over time. More important than the individual variable results, however, is the fit of the model to the data. The combination of high censoring/low hazards and the assumption that all observations will eventually "fail" results in a predicted median survival time for this

---

[11] Here we examine only the time to the first event, even though in some cases more than one response occurs. We address the mode general issue of models for repeated events in a different context, below.

[12] A thorough discussion of these covariates and their expected effects can be found in Zorn and Caldeira (1995).

model of nearly 710 years (!), clearly a suboptimal fit to the data.

In contrast, the split-population model presented in columns 2 and 3 presents a somewhat different picture of the Congressional response data. The first column indicates the probability of a case being essentially “immune”; i.e., of its never being addressed by Congress, while the second shows the effects of the covariates on the (log of the) duration until such a response occurs, given that the case is among those for which a response is possible. The results are revealing: in most instances, we expect (and find) that the signs in the two parts of the model will be the same (i.e., variables which decrease the probability of a response also serve to increase the duration until such a response is forthcoming). So, for example, the presence of *amicus curiae* briefs both significantly increases the probability of a Congressional response, and decreases the length of time until that response occurs. At the same time, other variables appear to work at cross-purposes: liberal decisions by the Court, for example, are both less likely to be addressed , but also see responses more rapidly than do conservative cases, when they occur. Likewise, more recent decisions are both more likely to go ignored by Congress, but are also addressed more rapidly when such responses occur, than are older cases, though this result is likely due to older cases having greater “exposure” to response than more recent decisions.

Also important is the improvement in model fit gained by the split-population model. While the models are not amenable to standard likelihood-ratio tests, we note that, in contrast to the standard model, the split-population model predicts that the average long-term probability of a response is 0.675, while the median predicted survival time is reduced by 50 percent (to 355 years). While still imperfect by any standard, it is clear that the split-population model fits the data better than the standard model. The differences in these models are illustrated in Figure 3, which plots the predicted forty-year survival probabilities for a “median” case[13] for each of the two models. Conditional on a case being in the “non-immune” population, the estimated survival rates are significantly less than for the general model, suggesting that separation of likely from unlikely cases for response provides estimates which yield better leverage on the long-term probability of Congressional action.

In summary, split-population models offer the potential for substantial improvements in the manner in which political scientists study duration data. In many cases in the social sciences, it is unrealistic to believe that all observations will eventually experience the event of interest. In addition to the examples given here, others examples might include studies of international conflict (e.g. Beck et. al. 1998), policy

[13] That is, a 1972 decision with one amicus brief and zeros on all other independent variables.

diffusion (e.g. Mintrom 1997), and other areas of political science where duration models have already been widely used.

3. General Models for Unobserved Heterogeneity

In this section, we discuss two broad classes of models that have recently emerged for dealing with the general issue of unobserved heterogeneity in survival data and a sketch of these models is provided in Figure 4. We begin by considering Cox's (1972) proportional hazards model:

$$h(t) = \lambda_0(t) exp(\mathbf{X}_i\beta) \tag{3.1}$$

where λ_0 is an unspecified baseline hazard. Cox's approach is by far the most widely used survival model in medicine and biostatistics, and has received some application in political science as well (e.g. Katz and Sala 1996; Kessler and Krehbiel 1996; Box-Steffensmeier et. al. 1997; Pellegrini and Grant 1999). For the case of unobserved heterogeneity, we consider the model in the presence of unmeasured, individual-level effects:

$$h(t) = \lambda_0(t) exp(\mathbf{X}_i\beta + \alpha_i) \tag{3.2}$$

Both modeling approaches address heterogeneity due to omitted covariates or individual effects, and both are applicable to a wide range of circumstances from which such heterogeneity may arise, though the manners in which they do so differ considerably. *Variance-corrected models*, also known generally as *"marginal"* models, take advantage of the fact that, in the presence of misspecification, the standard Cox model estimates for β converge to a well-defined constant vector (usually termed β^*) which can often be interpreted meaningfully (Struthers and Kalbfleisch 1984; Lin and Wei 1989), but the estimated covariance matrix is inappropriate for hypothesis testing. These models then simply estimate a standard Cox model and adjust the variance-covariance matrix to account for the individual- or group-specific effects that remain. By contrast, *conditional* (or "*frailty*" or *mixture*) models explicitly model the α_i's, occasionally as fixed-, but more often as random-effects following a specific parametric distribution.[14] Random effects

[14] The fixed-effects approach, treats μ as a fixed variable, and is rarely used. While an advantage of this approach is that few assumptions are imposed on μ, the disadvantages outweight the advantages (see Andersen, Klein, and Zhang 1999).

models take into account the correlation among failure times as a random effect term.[15]

We illustrate these models via a consideration of repeated events data. An obvious and important source of heterogeneity in survival models arises from correlation among repeated events experienced by the same subjects. Crouchley and Pickles (1995) point out that such correlation is often a consequence from clustered sampling in surveys, or in multiple events data, which occur when there are two or more failures for some observations in the data set. Multiple (or multivariate) events data are referred to more specifically as repeated events, when there are identical events for related observations, or as competing events when there are different events for different observations in the data set. The occurrence of coups would be a repeated events example, while the timing of when two Senators from the same state announce a position on a major bill illustrates identical events for related observations. The study of congressional careers, by contrast, is an example of competing events, i.e., retire, defeat in the general election, run for higher office (Jones 1994). It is important to distinguish repeated events and identical events for related observations from competing events to account for the fact that the failure times are correlated (here, within nations and by states, respectively). This correlation violates the assumption that the timing of events are independent, and results in the problems of estimation and inference discussed above.

Therneau (1997) and Cleves (1999) provide helpful suggestions when analyzing a data set with multiple events. First, we should ask, are the events are ordered or unordered? Second, are they the same type of event, or different? Unordered events of different types call for a competing risks analysis. In this case there are several possible events that may occur for an observation in random order and each event may only occur once per observation.[16] In these circumstances, the baseline hazard can also be allowed to vary by event type. Unordered events of the same type may also be correlated: in the example of position taking by the two Senators from the same state, both senators may be asked for the first time about their position for a news story in a leading state paper, or one may cue off the other in making his or her decision. Such dynamics may lead to correlation among the timing of position statements by the two senators, which must be accounted for if proper inferences are to be made. In their work on this topic,

[15] Readers familiar with models for time-series cross sectional data will note the similarity between these approaches and "population averaged" and "subject-specific" models for panel data (e.g. Hu et. al. 1998).

[16] Lunn and McNeil (1991) point out that correlation does not seem to be a problem when each subject can contribute at most one event.

Arnold, Box-Steffensmeier, and Zorn (1998) do so by estimating robust standard errors, allowing heterogeneity within a state.

The analysis of ordered events is a more controversial and active area of research, and the focus of our discussion here. We consider three variance corrected models: the "independent increments" (AG) model developed by Anderson and Gill (1982), marginal risk set model of Wei, Lin, and Weissfeld (1989), and conditional risk set model (Prentice, Williams, and Peterson 1981), which may be estimated in either elapsed or interevent time. We also address the general approach of random effects, which accounts for unmeasured effects by assuming the random effects follow a specific distribution. Since the variance must be nonnegative, the most commonly-used distributions are the gamma (with mean one and unknown variance), Gaussian (normal) distribution, and *t* distribution. In the discussion that follows, we estimate and compare these models, using data on the incidence of militarized interstate disputes collected by John Oneal and Bruce Russett (1997).

3.1 Variance-Corrected Models

As noted above, variance-corrected models are fit as though the data were independent observations, and then the variance is then "fixed" to account for the interdependencies brought on by unobserved or omitted factors. Cleves notes that "the principal difference between these methods is in the way that the risk sets are defined at each failure" (1999, 34). The earliest variance-corrected model was that of Anderson and Gill (1982) (hereinafter AG), and is based on the counting process formulation of the Cox model (see, generally, Fleming and Harrington 1991). The key characteristic of the AG model is "the assumption that the risk of an event for a given subject is unaffected by any earlier events that occurred to the same subject, unless terms that capture such dependence are included explicitly in the model as covariates" (Oakes 1992, 372). That is, multiple events for any particular observation are independent;[17] for this reason, the AG model is often referred to as the "independent increment" model. If events are not independent, robust variance estimates allowing for clustering within units may be used (Hamilton and Therneau 1997, 2034).[18] They note that "(E)ffects that may change with event number may be modeled explicitly, using time dependent covariates. For instance, if **Z** is the time dependent covariate 'number of

[17] More specifically, that event arrivals follow an independent Poisson process.

[18] Note that if the independent increments assumption holds, the naive and robust standard errors will be equal.

previous events', one's model might include both treatment (effects), **Z** and their interaction" (Ibid.).

In practical terms, the Cox and AG models are similar so as in most cases to be indistinguishable, and in fact the former is a special case of the latter. Thus, the AG approach is straightforward to estimate, but the assumption if independent increments is strong, particularly if the ordering of events is significant. By contrast, the marginal risk set approach of Wei, Lin, and Weissfeld (1989) treats ordered events data as if it were a competing risks problem: each observation is "at risk" for the first, second, etc. event from the beginning of the study period. Their approach is thus referred to as the "marginal risk set" model because marginal data are used, "that is, 'what would result if the data recorder ignored all information except the given event type?'" (Therneau and Hamilton 1997, 2035). As a result, at any point in time, all observations who have not experienced k events are assumed to be "at risk" for the kth event. Estimates are then stratified for each event rank, which allows the baseline hazard rate to vary for each subsequent event,[19] but the effects of covariates are assumed to be constant across event ranks.[20]

The signature characteristic of the marginal approach is that all observations are at risk for all events at all times prior to experiencing that event. By comparison, in the conditional model of Prentice, et. al. (1981), an observation is not at risk for a later event until all prior events have occurred. Accordingly, the "risk set" at time t for the kth occurrence of an event is limited to those observations under study at t who have already experienced k-1 events of that type.[21] As in the marginal model, estimates are then stratified by event rank, so that the different event ranks may have varying baseline hazards, but covariate effects are assumed to be constant across strata. Oakes (1992) notes that, in the case of ordered events, the conditional model's accounting for event ordering results in efficiency gains over the marginal model. An additional feature of the conditional risks model, as introduced by Prentice et. al. (1981), is that the model may be estimated in either elapsed time (i.e., time from entry into the observation set) or in time from the

[19] Recall that stratification is used to permit flexible variation in the baseline hazard, not for an estimate of variable effects.

[20] One can, as Wei et. al. (1989) did, include strata-by-covariate interactions to estimate separate effects. Alternatively, one can estimate models for each strata separately; these separate estimates and standard errors by strata are algebraically equivalent to those which would result from a combined fit over all strata.

[21] Therneau and Hamilton (1997) clearly illustrate the difference between the marginal and conditional model by pointing out that if events occurred at 100 and 185 days and the subject has been observed for 250 days, then the marginal model treats the subject as "at risk" for the second event from day 0 to day 185. In contrast, since an observation cannot be at risk for event two until event one occurs, the conditional model only places the subject "at risk" for the former from day 100 to day 185.

previous event (where the "clock starts over" after the occurrence of an event). The latter correspond to "renewal" or "semi-Markov" models (e.g. Lancaster 1990, 85-97), of which the Poisson process is a special case.

All three of these models use robust variance estimates to account for unobserved, or at least unaccounted for, heterogeneity. Robust standard errors are based on the idea that observations are independent across groups or clusters but not necessarily within groups. The robust variance estimator is then based on a 'sandwich' estimate:

$$\mathbf{V} = \mathbf{I}^{-1}\,\mathbf{B}\,\mathbf{I}^{-1} \tag{3.3}$$

where $\mathbf{I}^{-1}$ is the usual variance estimate of a Cox model (the inverse of the information matrix $\mathbf{I}$) and $\mathbf{B}$ is a correction factor. There are several ways to motivate this correction: as the proper variance when a likelihood for distribution f is fit, but the data come from g (Huber 1967); as the multivariate form of a variance inflation factor (Smith and Hietjan forthcoming); or as the approximation to the jackknife estimate (Therneau 1997). Because the unobserved intracase correlations are generally positive, the estimates of the variance-corrected standard errors are almost always larger than those from a "naive" estimate based in $\mathbf{I}^{-1}$.

Considerable care needs to be taken when setting up the data for the variance corrected models, particularly in specifying which observations are considered "at risk" for the event in question. Therneau (1997), Therneau and Hamilton (1997), and Cleves (1999) present exceptionally clear descriptions of this process. For the AG and conditional models, each subject is represented as a set of rows with time intervals of (entry time, first event time], (first event time, second event time], . . . (kth event time, final observation time]. A case with no events would thus have a single (censored) observation; a case with one event occurrence, and that did not end at the last time of observation, would have two events; etc. For the marginal model, each observation appears in the data once for each possible event rank. So, if the researcher observed a maximum of k events, each subject would appear in the data k times, once for each event rank. For all three variance corrected models, time-varying covariates may be included, though at what may be a significant complication to the data setup.

3.2 Frailty Models

An alternative set of approaches for dealing with heterogeneity are what are (somewhat unfortunately) referred to as "conditional" models, and fall into two general categories. *Fixed effects* models deal with the α_i's in (3.2) by explicitly including those unit effects in the model estimation, and estimating their impact. Such approaches have not been widely discussed or used in the context of survival modeling (a notable exception is Allison 1996), and are generally not favored as an approach for dealing with heterogeneity.[22] By comparison, *frailty (*or random-effects) models typically consider the model:

$$h(t) = \lambda_0(t)\mu_i \, exp(\mathbf{X}_i\beta) \qquad \textbf{(3.4)}$$

where the μ_i are equal to $exp(\alpha_i)$ from (3.2). This formulation is often referred to as the *multiplicative frailty* model, because the individual-specific effects (or "frailties") operate in a multiplicative fashion on the baseline hazard (e.g. Vaupel et. al. 1979; Aalen 1994; Keiding et. al. 1997).[23]

The underlying logic of these models is that some individuals (or groups or clusters) are intrinsically more or less prone to experiencing the event of interest than are others, and that the distribution of these individual-specific effects can be known, or at least approximated. Accordingly, a specific parametric distribution is often assumed for the μ's. Because the hazards are necessarily positive, the distribution of μ's is usually chosen from the class of positive distributions; in applied work, the most widely used are the gamma, normal, and *t* distributions, with the gamma being by far the most frequent.[24] Conditional on the chosen parametric distribution, the event times are assumed to be independent, so inference may be made in standard fashion, though robust standard errors (White 1980) are still used.

Estimation is accomplished by first deriving the likelihood for the observed history on an individual, conditional on observed and unobserved variables. Then one uses the imposed distribution of the unobservable to compute the mean of the likelihood when the unobserved are not taken into account.

---

[22] Andersen et. al., for example, found that their random-effects score test performed significantly better than score, Wald or likelihood-ratio tests based on fixed effects in assessing the presence of cluster-specific effects in the context of the proportional hazards model.

[23] Oakes (1992) states clearly that "a frailty is an unobserved random proportionality factor which applies to the hazard function for each subject" (1992, 372).

[24] One reason for the gamma's predominance is its analytical tractability: the connection between Cox's model and a homogenous Poisson model with time-specific effects (Laird and Oliver 1981) may be extended to the case of gamma heterogeneity, such that estimation may be accomplished with software capable of estimating negative binomial regression models (see e.g. Lawless 1987; Thall 1988; Abu-Libdeh et. al. 1990; Lindsey 1995).

This procedure is repeated for all individuals in the sample and then maximized.[25] Estimation yields parameter values and standard errors, as well as an estimate of the variance of the frailty distribution, the latter of which may be used to test the null hypothesis of no individual effects (i.e., independence of events within observations or groups).

3.3 Model Comparisons and Results

Thus, while the variance corrected models treat the dependence of related failure times as a nuisance, frailty models explicitly formulate the nature of dependence (Lin 1994, 2246). That is, while variance corrected models rely on the consistency of the parameter estimates and then incorporate the effect of dependence among the repeated events by adjusting the standard errors, frailty models condition out the individual-specific effects to make accurate inferences. Lin (1994) points out that there has been considerable controversy over whether the variance corrected or random effects approach "is more naturally related to the underlying mechanisms. The latter approach is expected to be more efficient than the former provided that the frailty distribution is correctly specified" (Lin 1994, 2246).[26] In fact, however, both approaches have distinct advantages and disadvantages.

With respect to the variance corrected models, Therneau (1997) and Therneau and Hamilton (1997) show that each of the variance corrected models has potential biases in a study of hidden covariate effects.[27] Hidden covariate simulations mimic the case of important unmeasured or omitted covariates, a central motivation for models with heterogeneity. Based on simulations of the variance corrected models with a missing covariate, Therneau (1997) concludes that the AG model is most reliable for overall estimates of covariate effects. Lin (1994) also shows that the marginal model, but not the conditional model, is robust for testing the overall effect of an independent variable. The robust estimate of the variance should always be used when there are correlated event times. The AG model does not allow one to investigate effects that might change based on event number except via time-dependent covariates. That is, the AG model only yields overall effect for the independent variables when, in fact one may want to know the effects of the independent variables for the second, third, or subsequent events as well. Lin goes on to state that specifically constructed time-dependent covariates for capturing such effects may be

[25] See Lancaster (1979) for elaboration.

[26] We revisit the issue of the specification of the random effects distribution below.

[27] See Oakes (1992), Omori and Johnson (1993), and Lin (1994) as well.

misleading (1994).

In regard to the marginal model, a number of concerns have been raised. Cook and Lawless (1997) question the assumptions of the marginal model with respect to the specification of the risk set; relatedly, Oakes (1992; 1997) notes that, relative to the conditional model, the marginal model is inefficient. An equally important issue is that the marginal model may badly strain the proportional hazards assumption (Therneau 1997). Proportional hazards are a central premise of the Cox model and its variants, and one should always assess the validity of assuming proportional hazards, irrespective of the model (Box-Steffensmeier and Zorn 1998), but analysts should be aware that the violation is more likely in the marginal model than in the AG model. Conversely, "The marginal model has the advantage of giving estimates of the possible change in treatment effect over time" through the use of strata-by-variable interactions (Therneau and Hamilton 1997, 2044). This suggests that applied researchers may benefit from the use of both models. If the coefficients of the AG and marginal models are similar, and the proportional hazards assumption is not violated, one is on good footing to both draw inferences about variable effects, and to conclude that there is not a missing covariate that is distorting the coefficients.

Intuitively, the conditional model's preservation of the order of sequential events in the creation of the risk set renders it an attractive alternative. Paradoxically, however, Therneau and Grambsch (1998) show that the this characteristic of the conditional model means is severely limited by its sensitivity to loss of randomization across strata (i.e., events) in clinical trials. The intuition of this critique is that, to be at risk in higher strata (e.g., for later events), one must have experienced the event in question. To the extent that treatment effects are present and influential on the hazard of the event, observations at risk for the second and subsequent events present a nonrandom sample of the data, biasing parameter estimates and making inference difficult. Moreover, for higher ranked events, the risk set may be very small (since few observations have experienced the events necessary to place them in the risk set for that event), yielding estimates that may be both unstable and imprecise.

Random effects models are subject to two primary criticisms, both relating to the specification of the random effects distribution. The first concerns the choice if the distribution itself. Neither theory nor data typically provides much guidance for imposing a specific distribution on the frailties, and "parameter estimates *can* be highly sensitive to the assumed parametric form of the error term" (Blossfeld and Rohwer 1995, 255) . Heckman and Singer (1982, 1984b, 1985) criticize the parametric assumptions of previous frailty models and develop an estimator of the hazard rate that is semi-parametric with respect to the distribution of the individual-specific effects. It is important to note, however, that their estimator is itself sensitive to the parametric form of the hazard chosen for the general model, and to the number and choice of explanatory variables.[28] There is continuing research on how to best choose a distribution (Schumacher, Olschewski, and Schmoor 1987; Lancaster 1990; Hougaard 1991; Vaupel and Yashin 1995; Sastry 1997). Among the most promising is that of Commenge and Andersen (1995), who develop a test for random effects that does not require specification of the unknown error term distribution (see also Andersen et, al. 1999).

A second, related point of criticism is that the random effects distribution is required to be independent of the model's included covariates. As noted by Vermunt (1997) and others, this assumption flies in the face of the omitted-variables justification often given for random-effects models: "(I)f one assumes that particular important variables are not included in the model, it is usually implausible to assume that they are completely unrelated to the observed factors" (1997, 196). Chamberlain (1985) points to this difficulty as a central motivating reason for the use of fixed effects models, though those models are also subject to their own problems.

Conversely, an advantage of random effects models is their ability to analyze data that are also correlated at several different hierarchical levels (Goldstein 1995; Bandeen-Roche and Liang 1996; Sastry 1997). For example, if we are studying survey respondent's attitudes toward government corn subsidies, we may want to cluster by state and region. In internal conflict studies by dyad, we may want to cluster by country and by region. Clustering by country would be important because some countries may be more prone to internal conflict, due to unmeasured or unmeasurable factors. Clustering by region could address a controversy in the literature about whether or not internal conflicts diffuse across boarders (contagion) or escalate to bring in other powers (e.g., the U.S. in Kosovo) (Sweeney 1999).

We compare variance corrected and random effect approaches to heterogeneity, using Oneal and Russett's (1997) data on the relationship among economic interdependence, democracy, and peace.[29] Data consist of 20990 observations on 827 "politically relevant" dyads between 1950 and 1985. We model the hazard of a militarized international conflict as a function of six primary covariates: a score for *democracy* (based on the Polity III data), the level of *economic growth*, the presence of an *alliance* in the dyad, the two nation's *contiguity*, their military *capability ratio*, and the extent of bilateral *trade* in the dyad.[30] Liberal

[28] See Trussel and Richards (1985), Kiefer (1988), and Hoem (1990).

[29] Beck et. al. (1998) also use this data, which is available at ftp://weber.ucsd.edu:/pub/nbeck.

[30] See Oneal and Russett (1997) for details of the variables and coding.

theory suggests that all variables except contiguity ought to decrease the hazard of a dispute, while contiguity should increase it. Here, following Beck et. al. (1998), we limit our analysis to observations which are not continuations of conflicts, yielding a valid N of 20448 observations. We address the issue of heterogeneity in two ways: by considering cross-sectional heterogeneity by dyad, and (more important) by addressing the matter of repeated events and dependency within dyads.[31]

We begin by examining the results of our estimation of the various marginal models, presented in Table 4. As a point of comparison on the question of multiple events, we include a model which only considers the time to each dyad's first post-War dispute, omitting second and subsequent disputes entirely, as well as a standard Cox model of all disputes. One obvious similarity for all the models is the result that each model is preferred to the null model of no independent variables, as indicated by the likelihood ratio (LR) or Wald tests.[32] The model for first events uses data only on the time until the first event, and thus implicitly assumes that the first event is representative of all events, a questionable assumption here as in most situations.[33] In addition, omitting second and subsequent observations results in a loss of information. The potential gain in efficiency from examining the additional data is equal to the square root of the number of first events divided by the number of total events; here, that figure is equal to 0.92, or 8 percent. That efficiency is reflected in the estimates for the standard Cox proportional hazards model in column two: the standard errors are uniformly smaller for the Cox model, which uses all observations, than for the model of first events, though only the estimate for the trade variable differs dramatically between the two models.[34] More generally, if one's theory suggests the *k*th event is different from the *k*-1th, then a model based only on first events is inappropriate.

---

[31] A similar effort that also addresses the question of left censoring in Beck et. al.'s (1998) *Peace Years* variable is Reiter (1999).

[32] The LR test is used for the first two models in Table 4, where the data are assumed to be uncorrelated; the Wald test is used for the remaining five models. Efron's approximation for ties is used for all models. The Breslow approximation was the first approximation developed and is not generally recommended, while the exact likelihood option in all cases gave results extremely similar to those obtained with the Efron approximation (typically within the thousandths decimal place). The exact likelihood method usually takes considerably longer to converge.

[33] The exception is, of course, when one is intrinsically interested in modeling the time to a first event; e.g., entry of the first high quality challenger in an incumbent's reelection race.

[34] See Beck et. al. (1998) for a discussion of the connection between repeated events and the effect of trade on international conflict.

The third model in Table 4 improves upon the Cox model by taking into consideration the fact that the same dyads are contributing more than one observation. The Anderson-Gill (AG) model gives the same parameter estimates as the Cox model, but estimates robust standard errors clustered by dyad to account for repeated measures. Here, the standard errors are larger for the AG than for the Cox model: the positive correlation across observations of a single dyad yields less information that if one considers all 20,448 observations as independent.[35,36]

As noted above, a signature characteristic of the AG model is its characterization of the underlying process generating the events as a homogenous Poisson process. This means that event- or duration-dependence in the AG model must be modeled explicitly, in the form of covariates. If theory suggests that your baseline hazards will be different for the *k*th event than for the *k*-1th event, a simple approach to accounting for this dependence is to include some function of past events or durations in the specification of the AG model. In the case of international conflict examined here, it is certainly reasonable to expect and test whether the history of peace affects future spells of peace. To do so, we include a simple counter for the number of previous disputes in that dyad since World War II. The result is generally to attenuate the coefficient estimates, with the exception of the variable for dyadic trade, which becomes positive and significant (again, see Beck et. al. 1998). However, adding the counter does not allow the hazard rates to vary by event number, which is an advantage of the marginal and conditional models.

Results for the marginal model show that the AG and marginal results are very similar, suggesting that treating the events as independent and clustering on dyad is an appropriate strategy.[37] The robustness of the results across these models is comforting, and conforms to Blossfeld and Rohwer's (1995) modeling strategy suggestions. They argue that social science theory is almost always too weak to suggest a particular model, and that instead one should aim for estimation results that are to a large degree

[35] It is possible that the robust standard errors will be smaller than the regular standard errors. Lee, Wei, and Amato (1992) present such a case and explain that, when one of the independent variables is balanced within subjects, there is an improvement in the robust standard errors, analogous to a paired *t*-test.

[36] Large differences between the conventional and robust standard errors would be evidence of an assumption violation for the Cox model (Therneau and Hamilton 1997, 2039).

[37] Graphical tests of the proportional hazards assumption (e.g. Grambsch and Therneau 1994), which are not shown here, suggest that the assumption is generally valid; however, it is also clear that hazards grow less proportional in the higher strata as the number of dyads experiencing a high number of events decreases.

independent of a specific model specification. If there were differences, and these differences were related to a plausible theory then more weight would be given to that model. For example, if after experiencing the first event the risk (hazard) of the next event increased, one would have more confidence in the marginal results, and may want to consider adjusting the AG model by incorporating a time varying covariate for the number of previous events interacted with the other independent variables.

The final two columns of Table 4 present the result of the conditional risks models, the first measuring time from entry and the second measuring time from previous event. These models yield results that are in some ways quite different from those already discussed: for example, the democracy, trade, and capability ratio variables are now all estimated as having a positive impact on the probability of a dispute, and those for the latter variable achieve statistical significance. A likely explanation for these differences is rooted in the differences between these models and those discussed above, in particular their differential specification of the risk set for a particular event. Because, in the conditional model, dyads are only at risk for a second dispute after they have had a first, a selection effect occurs: dyads with low scores for democracy or capability ratio enter into mode disputes, and thus move into the risk set for subsequent disputes, where they are the only ones considered when estimating the hazards of higher-ranked events. This selection yields estimates which are both difficult to interpret and also likely biased (Therneau 1999).

A useful way of comparing different Cox models is to examine similarities and differences in their predicted baseline hazards. Figure 5 plots smoothed baseline hazards for the four central models in Table 4. Several interesting characteristics are apparent. First, as we would expect, given the similarity of the coefficient estimates, the baseline hazards for the Andersen-Gill and marginal models are almost identical, and highly correlated at 0.92. In addition, both baseline hazards are very low, reflecting the fact that only a small number of observations at risk in these two models "fail". By contrast, the baseline hazards for the two conditional models are substantially higher, and also exhibit greater variability. The effect of the decreasing size of the conditional models' risk sets over time is also apparent, both in the higher overall predicted hazards and in the greater variability and instability of the estimates at later time points.

Contrasting with the variance corrected models are random-effects approaches to multiple events heterogeneity. We estimated the same six-variable model on the Oneal and Russett data, this time using a multiplicative random-effects specification and specifying three alternative distributions for the variance of the individual-specific components: the unit-mean gamma, Gaussian, and *t* distributions. In each case, the results were similar across all three models, suggesting that, for these data, the choice of the random effects distribution is of little consequence. In addition, all three models yielded estimates similar to those of the

AG and marginal models, a fact that bodes well in general for our estimates of the variable effects, as well as for other necessary assumptions of the models (e.g. proportional hazards).
Thus, while the choice of a distribution *can* be a problem, we see for our empirical data it is not; the estimation results are fairly robust across the three models.

Our random effects models also provide estimates of the variance of the distribution of random effects. Sastry explains that if the variance is zero, then observations from the same dyad are independent (1997, 430), and that a larger variance implies greater heterogeneity in frailty across dyads and greater correlation among events occurring within the same dyad (1997, 430). The variance of the random effect for all three error term distributions is different from zero to a statistically significant degree. At the same time, the standard error estimates for the frailty model coefficients are nearly identical to those found in the Cox and AG models. This, combined with the aforementioned similarity between the frailty and AG/marginal parameter estimates, suggests that, while dyad-specific effects are present in the data, they appear not to substantially effect estimation of the effects of our independent variables on the hazard of a dispute.

4. Conclusion

"We can usually be sure that we were not able to include all important covariates" (Blossfeld and Rohwer 1995, 243).[38] That note of pessimism ought to ring true to anyone engaged in serious empirical research in the social sciences. Accordingly, it is important in all empirical data sets to assess the potential biases that may arise from unobserved heterogeneity in one's data, and to employ methods appropriate for correcting those biases, if possible. Here, we have offered a survey of approaches for dealing with heterogeneity in the context of duration models. In closing, we consider some common threads running through these models, and address some recent research on providing further synthesis in this area.

As we note above, split-population models can be considered a special case of a model that mixes a standard survival distribution with a proportion of immune subjects; Therneau and Grambsch (1998) and Aalen (1992) highlight the connection of frailty models to split-population models when discussing variations in failure rates. Less appreciated is that the reverse is also true: frailty models in general are simply a broader class of the mixture model idea inherent in the split-population case. Vermunt (1997,

[38] More humorously, Therneau (1999, 258) offers a test for the null hypothesis (H_0) that all important covariates are included in the model. Test: Is it a simulation study? 1) Yes: Read the paper or ask the author if all the covariates are in there. 2) No: Reject H_0.

190) has noted that the multiplicative frailty model (3.4) “is, in fact, a mixture model ... (where) the mixture variable is assumed to have a particular continuous distribution function”. A distinctive characteristic of the split-population model, however, is that, unlike standard frailty models, the mixture distribution is not required to be independent of the covariates; in fact, the nature of the mixing distribution is specifically linked to the covariates of interest (Farewell 1982; Land et. al. 1996). Generalizing this approach offers a potential for solving the problematic independence assumption, and some initial steps in this direction have already been taken: Rohwer (1995), for example, includes in his *TDA* program an option to allow the variance of the gamma frailty distribution to depend on model covariates.

A related point of similarity lies in the intersection of semi-parametric duration models and log-linear models for event count processes (e.g. Laird and Oliver 1981; Lindsey 1995, 1998). Models for event counts have seen rapid development in recent years, and a number of those offer the potential for innovative application in the area of duration models. Among these are the variance-function negative binomial models for gamma-distributed heterogeneity mentioned previously, as well as marginal models based on an estimating equations approach to inference in the presence of clustered data (e.g. Diggle et. al. 1994). Regarding the latter, Therneau (1999) notes that the variance-correction approach to multiple events data is precisely equivalent to the “working independence” correlation structure widely used in GEE models (Liang and Zeger 1986). Xue (1998) offers an initial foray into this area as well.

A final area of research involves uniting variance corrected and frailty models explicitly. Recent work on conditional models, for example, has attempted to rectify the difficulty with selection into the risk set by considering a conditional model with random effects. Similarly, Price (1999) proposes a model that unites multiplicative frailty with a split population model so that there is a cured proportion. Tsodikov (1998) has offered a proportional hazards model that allows for a cured fraction, and Tsidokov et. al. (1999) extend that model to the case of time-varying covariates. Finally, a number of recent studies have considered the effects of missing data on frailty models (Turnbull et. al. 1997; Jiang et. al. 1999).

References

Aalen, Odd O. 1988. "Heterogeneity in Survival Analysis." *Statistics in Medicine* 7:1121-37.

Aalen, Odd O. 1992. "Modeling Heterogeneity in Survival Analysis by the Compound Poisson Distribution." *The Annals of Applied Probability* 2:951-72.

Aalen, Odd O. 1994. "Effects of Frailty in Survival Analysis." *Statistical Methods in Medical Research* 3:227-43.

Abu-Libdeh, Hasan, Bruce W. Turnbull and Larry C. Clark. 1990. "Analysis of Multi-Type Recurrent Events in Longitudinal Studies; Application to a Skin Cancer Prevention Trial." *Biometrics* 46(December):1017-34.

Allison, Paul D. 1996. "Fixed Effects Partial Likelihood for Repeated Events." *Sociological Methods and Research* 24:207-22.

Andersen, Per Kragh and R. D. Gill. 1982. "Cox's Regression Model for Counting Processes: A Large Sample Study." *The Annals of Statistics* 10(4):1100-20.

Andersen, Per Kragh, John P. Klein and Mei-Jie Zhang. 1999. "Testing for Centre Effects in Multi-Centre Survival Studies: A Monte Carlo Comparison of Fixed and Random Effects Tests." *Statistics in Medicine* 18:1489-1500.

Arnold, Laura W., Janet M. Box-Steffensmeier and Christopher J. W. Zorn. 1998. "Timing and Position Taking in the Senate." Paper presented at the Annual Meeting of the American Political Science Association, September 3-6, 1998, Boston, MA.

Bandeen-Roche, Karen J., and Kung-Yee Liang. 1996. "Modelling Failure-Time Associations in Data with Multiple Levels of Clustering." *Biometrika* 83: 29-39.

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(October):1260-88.

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

Berkson, J. and R. P. Gage. 1952. "Survival Curve for Cancer Patients Following Treatment." *Journal of the American Statistical Association* 47:501-15.

Boag, J. W. 1949. "Maximum Likelihood Estimates of the Proportion of Patients Cured by Cancer Therapy." *Journal of the Royal Statistical Society, Series B* 11(1):15-44.

Box-Steffensmsier, 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):324-38.

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

Box-Steffensmeier, Janet M. and Peter M. Radcliffe. 1996. “The Timing of PAC Contributions: A Split-Population Survival Model.” Paper presented at the Annual Meeting of the American Political Science Association, San Francisco, CA.

Box-Steffensmeier, Janet M. and Christopher J. W. Zorn. 1998. “Duration Models and Proportional Hazards in Political Science.” Paper presented at the Annual Meeting of the Midwest Political Science Association, April 23-25, 1998, Chicago, IL.

Blossfeld, Hans-Peter and Götz Rohwer. 1995. *Techniques of Event History Modeling.* Mahwah, NJ: Lawrence Erlbaum.

Cai, Jianwen, Pranab Sen, and Haibo Zhou. 1999. "A Random Effects Model for Multivariate Failure Data." *Biometrics* 55:182-9.

Chamberlain, G. 1985. “Heterogeneity, Omitted Variable Bias, and Duration Dependence.” In J. J. Heckman and B. Singer (eds.), *Longitudinal Analysis of Labor Market Data*. Cambridge, UK: Cambridge University Press.

Clayton, David. 1999. “Poisson Regression with a Random Effect.” *Stata Technical Bulletin* 46:30-33.

Cleves, Mario. 1999. “Analysis of Multiple Failure-Time Data with Stata.” *Stata Technical Bulletin* 49:30-39.

Commenges, D., and P. K. Andersen. 1995. “Score Test of Homogeneity for Survival Data.” *Lifetime Data Analysis*. 1: 145-60.

Cook, Richard J. 1995. “The Design and Analysis of Randomized Trials with Recurrent Events.” *Statistics in Medicine* 14:2081-98.

Cook, Richard J. and J. F. Lawless. 1997. "An Overview of Statistical Methods for Multiple Failure Time Data in Clinical Trials - Discussion." *Statistics in Medicine* 16(8):841-43.

Dingle, Peter J., Kung-Yee Lang., and Scott L. Leger. 1994. *Analysis of Longitudinal Data*. New York: Oxford University Press.

Eskridge, William N., Jr. 1991. "Overriding Supreme Court Statutory Interpretation Decisions." *The Yale Law Journal* 101:331-455.

Farewell, V. T. 1982. “The Use of Mixture Models for the Analysis of Survival Data with Long-Term Survivors.” *Biometrics* 38:1041-46.

Goldstein, Harvey. (1995). Multilevel Statistical Models London, Edward Arnold: New York, Halstead Press.

Greene, William H. 1997. *Econometric Analysis*, 3rd Ed. Upper Saddle River, NJ: Prentice-Hall.

Hettinger, Virginia and Christopher J. W. Zorn. 1999. "Signals, Models, and Congressional Overrides of the Supreme Court." Paper presented at the Annual Meeting of the Midwest Political Science Association, April 15-17, 1999, Chicago, IL.

Heckman, James and B. Singer. 1982. "Population Heterogeneity in Demographic Models." In K. Land and A. Rogers (e's.), *Multidimensional Mathematical Demography*. Cambridge, UK: Cambridge University Press.

Heckman, James and B. Singer. 1984. "The Identifiability of the Proportional Hazards Model." *Review of Economic Studies* 51:234-41.

Hougaard P. 1984. "Life Table Methods for Heterogeneous Populations: Distributions Describing the Heterogeneity." *Biometrika* 71: 75-83.

Hougaard P. 1986a. "Survival Models for Heterogeneous Populations Derived from Stable Distributions." *Biometrika* 73: 387-96.

Hougaard P. 1986b. "A Class of Multivariate Failure Time Distributions." *Biometrika* 73: 671-8.

Hougaard, P. 1991. "Modeling Heterogeneity in Survival Data." *Journal of Applied Probability* 28(3): 695-70.

Hougaard P. 1999. "Fundamentals of Survival Data." *Biometrics* 55(March):13-22.

Hu, Frank B., Jack Goldberg, Donald Hedeker, Brian R. Flay and Mary Ann Pentz. 1998. "Comparison of Population-Averaged and Subject-Specific Approaches for Analyzing Repeated Binary Outcomes." *American Journal of Epidemiology* 147(7):694-703.

Jiang, Wenxin, Bruce W. Turnbull and Larry C. Clark. 1999. "Semiparametric Regression Models for Repeated Events with Random Effects and Measurement Error." *Journal of the American Statistical Association* 94(445):111-24.

Jones, Bradford S. 1994. A Longitudinal Perspective on Congressional Elections. Ph.D. diss. State University of New York at Stony Brook.

Katz, Jonathan N., and Brian R. Sala. 1996. "Careerism, Committee Assignments, and the Electoral Connection." *American Political Science Review* 90: 21-33.

Keiding, Niels, Per Kragh Andersen and John P. Klein. 1997. "The Role of Frailty Models and Accelerated Failure Time Models ion Describing Heterogeneity Due To Omitted Covariates." *Statistics in Medicine* 16:214-24.

Kessler, Daniel, and Keith Krehbiel. 1996. "Dynamics of Cosponsorship." *American Political Science Review* 90: 555-567.

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

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

Land, Kenneth C., Patricia L. McCall, and Daniel S. Nagin. 1996. "A Comparison of Poisson, Negative Binomial, and Semiparametric Mixed Poisson Regression Models With Empirical Applications to Criminal Careers Data." *Sociological Methods and Research* 24: 387-442.

Lawless, J. F. 1987. "Regression Methods for Poisson Process Data." *Journal of the American Statistical Association* 82(399):808-15.

Lawless, J. F. and C. Nadeau. 1995. "Some Simple Robust Methods for the Analysis of Recurrent Events." *Technometrics* 37(May):158-68.

Lang., Kung-Yee and Scott L. Leger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73(1):13-22.

Lin, D. Y. 1994. "Cox Regression Analysis of Multivariate Failure Time Data." *Statistics in Medicine* 15: 2233-47.

Lin, D. Y. and L. J. Wei. 1989. "The Robust Inference for the Cox Proportional Hazards Model." *Journal of the American Statistical Association* 84(408):1074-78.

Lindsey, J.K. 1995. "Fitting Parametric Counting-Processes by Using Log-linear Models." *Applied Statistics - Journal of the Royal Statistical Society, Series C* 44(2):201-12.

Lindsey, J.K. 1998. "Counts and Times to Events" *Statistics in Medicine* 17:1745-51.

Longini, I. M. and M. E. Halloran. 1996. "A Frailty Mixture Model for Estimating Vaccine Efficacy." *Applied Statistics* 45:165-73.

Lunn, A.D., and McNeil, D.R. 1991. *Computer-interactive Data Analysis*. New York: Wiley.

Maller, R. A. and S. Zhou. 1996. *Survival Analysis with Long-Term Survivors*. New York: Wiley.

Manton, K.G., E. Stallard, and J.W. Vaupel. 1981. "Methods for Comparing the Mortality Experience of Heterogeneous Populations. *Demography* 18: 389-410.

Oakes, D.A. 1992. "Frailty Models for Multiple Event Times." In J.P. Klein and P.K. Goel (e's.) *Survival Analysis, State of the Art*. Netherlands: Kluwer Academic Publishers.

Oakes, D. (1997). "Model-Based and/or Marginal Analysis for Multiple Event-Time data?" In Lin, D.Y. and Fleming, T.R. *Proceedings of the First Seattle Symposium in Biostatistics*, 85-98, Springer, New York.

Omori, Y., and R.A. Johnson. 1993. "The Influence of Random Effects on the Unconditional Hazard Rate and Survival Function." *Biometrika* 80: 910-24.

Oneal, John R., and Bruce Russett. 1997. "The Classical Liberals Were Right: Democracy, Interdependence, and Conflict, 1950-1985." *International Studies Quarterly*, 41: 267-94.

Pellegrini, Pasquale A. and J. Tobin Grant. 1999. "Policy Coalitions in the U.S. Congress: A spatial duration modeling approach." *Geographical Analysis*. Vol. 31, 1 (January), 45-66.

Pickles, Andrew, and Robert Crouchley. 1995. "A Comparison of Frailty Models for Multivariate Survival Data." *Statistics in Medicine* 14:1447-61.

Price, Dionne. 1999. "Survival Models for Heterogenous Populations With Cure." Manuscript: Rollins School of Public Health, Emory University.

Reiter, Dan. 1999. "Issues of Left Censoring and Multiple Failures in Event History Analysis: An Application to International Conflict." Manuscript: Emory University.

Sargent, Daniel J. 1998. "A General Framework for Random Effects Survival Analysis in the Cox Proportional Hazards Setting." *Biometrics* 54(December):1486-97.

Sastry, Naryan. 1997. "A Nested Frailty Model for Survival Data, With an Application to the Study of Child Survival in Northeast Brazil." *Journal of the American Statistical Association* 92(438): 426-35.

Schmidt, Peter and Anne D. Witte. 1989. "Predicting Recidivism Using 'Split-Population' Survival Time Models." *Journal of Econometrics* 40(1):141-59.

Smith, Renée M. N.D. "The Timing of Presidential Addresses." Typescript. University of Rochester.

Sweeney, Kevin. 1999. "Internal Conflict and Multi-level Modeling." Typescript. Ohio State University.

Thall, P. F. 1988. "Mixed Poisson Regression Models for Longitudinal Interval Count Data." *Biometrics* 44:197-209.

Therneau, Terry M. 1997. "Extending the Cox Model." *Proceedings of the First Seattle Symposium in Biostatistics*. New York: Springer-Verlag.

Therneau, Terry M. 1999. "Modern Survival Analysis in S-Plus." Seattle: MathSoft, Inc.

Therneau, Terry M., and Patricia M. Grambsch. 1998. "Penalized Cox Models and Frailty." Typescript. Mayo Clinic. http://www.mayo.edu/hsr/biostat.html, June 20, 1999.

Therneau, Terry M. and Scott A. Hamilton. 1997. "rhDNase as an Example of Recurrent Event Analysis." *Statistics in Medicine* 16(18): 2029-47.

Tsodikov, A. 1998. "A Proportional Hazards Model Taking Account of Long-Term Survivors." *Biometrics* 54:1508-15.

Tsodikov, A., M. Loeffler and A. Yakovlev. 1998. "A Cure Model with Time-Changing Risk Factor: An Application to the Analysis of Secondary Leukaemia." *Statistics in Medicine* 17:27-40.

Turnbull, Bruce W., Wenxin Jiang and Larry C. Clark. 1997. "Regression Models for Recurrent Event Data: Parametric Random Effects Models with Measurement Error." *Statistics in Medicine* 16:853-64.

Vaupel, J. W. 1990. "Relatives' Risk: Frailty Models of Life History Data." *Theoretical Population Biology* 37: 220-34.

Vaupel, J. W., and A. I. Yashin. 1985. "The Deviant Dynamics of Death in Heterogeneous Populations." In *Sociological Methodology* (E. F. Borgatta, ed.) 179-211. San Francisco: Jossey-Bass.

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

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

Wang, Mei-Cheng and Shu-Hui Chang. 1999. "Nonparametric Estimation of a Recurrent Survival Function." *Journal of the American Statistical Association* 94(445):146-53.

Wei, L. J. and David V. Glidden. 1997. "An Overview of Statistical Methods for Multiple Failure Time Data in Clinical Trials." *Statistics in Medicine* 16:833-39.

Wei, L. J., D. Y. Lin and L. Weissfeld. 1989. "Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions." *Journal of the American Statistical Association* 84(408):1065-73.

Werner, Suzanne. 1999. "The Precarious Nature of Peace: Resolving the Issues, Enforcing the Settlement, and Renegotiating the Terms." *American Journal of Political Science* 43(July):912-34.

Xue, Xiaonan. 1998. "Multivariate Survival Data Under Bivariate Frailty: An Estimating Equations Approach." *Biometrics* 54(December):1631-37.

Zorn, Christopher J. W. 1998. "An Analytic and Empirical Examination of Zero-Inflated and Hurdle Poisson Specifications." *Sociological Methods and Research* 26(February):368-400.

Zorn, Christopher and Gregory A. Caldeira. 1995. "Separation of Powers: Congress, the Court, and Interest Groups." Paper presented at the Annual Meeting of the Public Choice Society, Long Beach, CA, March 24-26, 1995.

Table 1. Split Population Survival Model – 1993-94 Corporate PACs

	Large PACs						Small PACs					
	Timing			Likelihood			Timing			Likelihood		
Variable	**Coef.**	**Z-Score**	**P-Value**	**Coef.**	**Z-Score**	**P-Value**	**Coef.**	**Z-Score**	**P-Value**	**Coef.**	**Z-Score**	**P-Value**
Constant	6.026	70.495	0.000	3.447	29.147	0.000	6.942	50.502	0.000	5.560	41.208	0.000
Candidate Power												
Energy	-0.029	-1.210	0.226	-0.748	-21.168	0.000	0.068	1.756	0.079	-0.580	-14.782	0.000
Prestige	-0.004	-0.204	0.839	-0.330	-13.531	0.000	-0.012	-0.422	0.673	-0.201	-7.068	0.000
Seniority	0.311	2.583	0.010	-0.597	-3.726	0.000	0.345	1.787	0.074	-0.995	-5.353	0.000
DLeader	-0.084	-2.799	0.005	-0.658	-15.743	0.000	-0.020	-0.431	0.667	-0.507	-10.802	0.000
RLeader	0.004	0.142	0.887	-0.521	-11.230	0.000	0.095	1.971	0.049	-0.435	-8.791	0.000
Candidate Ideology												
Republicans	-0.055	-2.080	0.038	0.191	5.211	0.000	-0.198	-4.534	0.000	0.390	8.989	0.000
COC	-0.076	-1.486	0.137	-1.444	-20.934	0.000	0.398	4.890	0.000	-1.639	-20.424	0.000
Candidate Need												
VotePct	0.126	1.777	0.076	0.621	6.190	0.000	-0.597	-5.384	0.000	0.385	3.508	0.001
Quality	0.061	2.414	0.016	-0.275	-7.869	0.000	0.092	2.101	0.036	-0.241	-5.644	0.000
PQuality	0.031	0.799	0.424	-0.112	-2.051	0.040	0.018	0.299	0.765	-0.245	-4.058	0.000
HHincome	0.266	2.324	0.020	-0.210	-1.339	0.181	-0.380	-2.127	0.033	-0.054	-0.300	0.764
PacShare	-0.580	-8.577	0.000	-1.936	-22.538	0.000	-0.616	-5.568	0.000	-1.173	-10.798	0.000
BCash	0.113	2.580	0.010	-0.144	-2.745	0.006	0.055	0.805	0.421	0.006	0.091	0.928
PAC Resources and Geography												
LagPRecp	-0.608	-12.722	0.000	-1.708	-21.004	0.000	-3.562	-7.938	0.000	-16.811	-35.480	0.000
RecpSqrd	0.117	5.984	0.000	0.059	1.168	0.243	8.275	5.718	0.000	43.190	25.620	0.000
StShare	-0.344	-10.101	0.000	-2.104	-33.227	0.000	-0.180	-5.131	0.000	-2.823	-69.240	0.000

Notes: Large Corporate PACs: N = 58,253; Estimated Split = .259; Observed Split = .238; Sigma = 0.504; -2 Log Likelihood = 87,527.14;
Small Corporate PACs: N = 289,142; Estimated Split = .037; Observed Split = .030; Sigma = .532; -2 Log Likelihood = 86,495.76.

Table 2

Predicted Probability of a Contribution

Type of PAC	Baseline	Power	Ideology	Need	Geography
Large Corporate	0.177	0.304	0.236	0.234	0.210
Small Corporate	0.025	0.044	0.037	0.031	0.033

Note: Cell entries are the probability that a PAC-candidate pair will experience a contribution.

Table 3

Standard and Split-Population Models of Congressional Responses to Supreme Court Decisions, 1979-1988

Variables	Standard Log-Logistic	Split-Population Model: Pr(No Response)	Split-Population Model: Duration to Response
(Constant)	15.827**	-8.074*	18.190**
	(1.109)	(3.645)	(1.612)
Year of Decision	-0.128**	0.105*	-0.169**
	(0.013)	(0.043)	(0.020)
Lower Court Disagreement	-0.349	-1.900*	0.300
	(0.203)	(0.779)	(0.321)
Formal Alteration of Precedent	-0.884	4.085*	-3.140
	(0.590)	(1.638)	(1.650)
Declaration of Unconstitutionality	1.002	-1.881	1.512
	(0.578)	(2.309)	(0.904)
Liberal Decision	-0.121	1.869**	-0.794*
	(0.206)	(0.591)	(0.332)
Number of *Amicus Curiae* Briefs	-0.083**	-0.169**	-0.054*
	(0.021)	(0.065)	(0.023)
Federal Government Loser	-0.230	-3.338	0.599
	(0.285)	(2.144)	(0.496)
State Government Loser	0.603	-2.639	1.509*
	(0.348)	(1.972)	(0.622)
Local Government Loser	-0.101	-6.059**	1.066
	(0.381)	(2.134)	(0.558)
Business Loser	-0.264	-1.250*	0.237
	(0.235)	(0.518)	(0.357)
Class Action Loser	-0.480	-1.315	-0.041
	(0.356)	(1.364)	(0.456)
Natural Person Loser	0.585	1.943**	-1.218
	(0.309)	(0.643)	(0.713)
α	1.099**	1.117**	
	(0.129)	(0.138)	
*ln*L	-730.31	-707.27	

Note: N = 7033. Entries are MLE's; standard errors are in parentheses. One asterisk indicates $p < .05$, two indicate $p < .01$ (two-tailed). See text for details.

Table 4

Variance Corrected Models for Repeated Events

	Time to First Event		Cox Proportional Hazards Model		Anderson-Gill		Anderson-Gill with Previous Events		Marginal Model		Conditional Risks: Time From Entry		Conditional Risks: Time From Previous Event	
	β (s.e.)	*p*-value	β (s.e.)	*p*-value	β (robust)	*p*-value	β (robust)	*p*-value	β (robust)	*p*-value	β (robust)	*p*-value	β (robust)	*p*-value
Democracy	-0.455 (0.132)	0.001	-0.439 (0.100)	<0.001	-0.439 (0.123)	<0.001	-0.333 (0.108)	0.002	-0.438 (0.123)	<0.001	0.162 (0.103)	0.115	0.099 (0.075)	0.19
Growth	-2.225 (1.723)	0.196	-3.227 (1.229)	0.009	-3.227 (1.318)	0.014	-2.702 (1.331)	0..042	-3.183 (1.302)	0.014	-3.766 (1.064)	<0.001	-3.422 (1.242)	0.006
Alliance	-0.420 (0.160)	0.009	-0.414 (0.111)	<0.001	-0.414 (0.170)	0.015	0.110 (0.114)	0.336	-0.409 (0.168)	0.015	0.144 (0.108)	0.182	-0.202 (0.094)	0.031
Contiguous	1.081 (0.170)	<0.001	1.213 (0.121)	<0.001	1.213 (0.178)	<0.001	0.448 (0.124)	<0.001	1.203 (0.176)	<0.001	0.287 (0.111)	0.010	0.618 (0.104)	<0.001
Capability Ratio	-0.192 (0.060)	0.001	-0.214 (0.051)	<0.001	-0.214 (0.082)	0.009	-0.162 (0.059)	0.006	-0.213 (0.081)	0.009	0.059 (0.029)	0.040	0.056 (0.025)	0.028
Trade	-3.183 (11.616)	0.784	-13.162 (10.327)	0.202	-13.162 (13.827)	0.341	11.487 (6.670)	0.085	-13.001 (13.717)	0.343	5.997 (6.504)	0.357	0.812 (9.604)	0.933
Previous Events	–	–	–	–	–	–	1.062 (0.078)	<0.001	–	–	–	–	–	–
Wald or LR test	122.06 (*df*=6)	<0.001	272.35 (*df*=6)	<0.001	92.92 (*df*=6)	<0.001	381.01 (*df*=7)	<0.001	93.20 (*df*=6)	<0.001	34.54 (*df*=6)	<0.001	51.09 (*df*=6)	<0.001
N	16991		20448		20448		20448		163584		20448		20448	

Table 5

Cox Frailty Models for Repeated Events Heterogeneity

	Cox with Gamma Frailty		Cox with Gaussian Frailty		Cox with *t* Frailty	
	β (robust)	*p*-value	β (robust)	*p*-value	β (robust)	*p*-value
Democracy	-0.365 (0.110)	0.005	-0.397 (0.108)	0.001	-0.441 (0.114)	<0.001
Growth	-3.684 (1.299)	0.006	-3.804 (1.280)	0.004	-3.921 (1.284)	0.003
Alliance	-0.370 (0.125)	0.028	-0.394 (0.123)	0.010	-0.446 (0.133)	<0.001
Contiguous	1.200 (0.167)	<0.001	1.153 (0.131)	<0.001	1.192 (0.140)	<0.001
Capability	-0.199 (0.055)	<0.001	-0.191 (0.050)	<0.001	-0.214 (0.060)	<0.001
Trade	-3.039 (10.308)	0.800	-5.278 (9.828)	0.620	-4.648 (9.748)	0.660
Variance of the Random Effect	2.42	<0.001	1.18	<0.001	1.46	<0.001
N	20448		20448		20448	

Note: Cell entries are coefficient estimates; numbers in parentheses are robust standard errors.

Figure 1

Estimated Transition Rate
Mixture of 2 Exponential Distributions

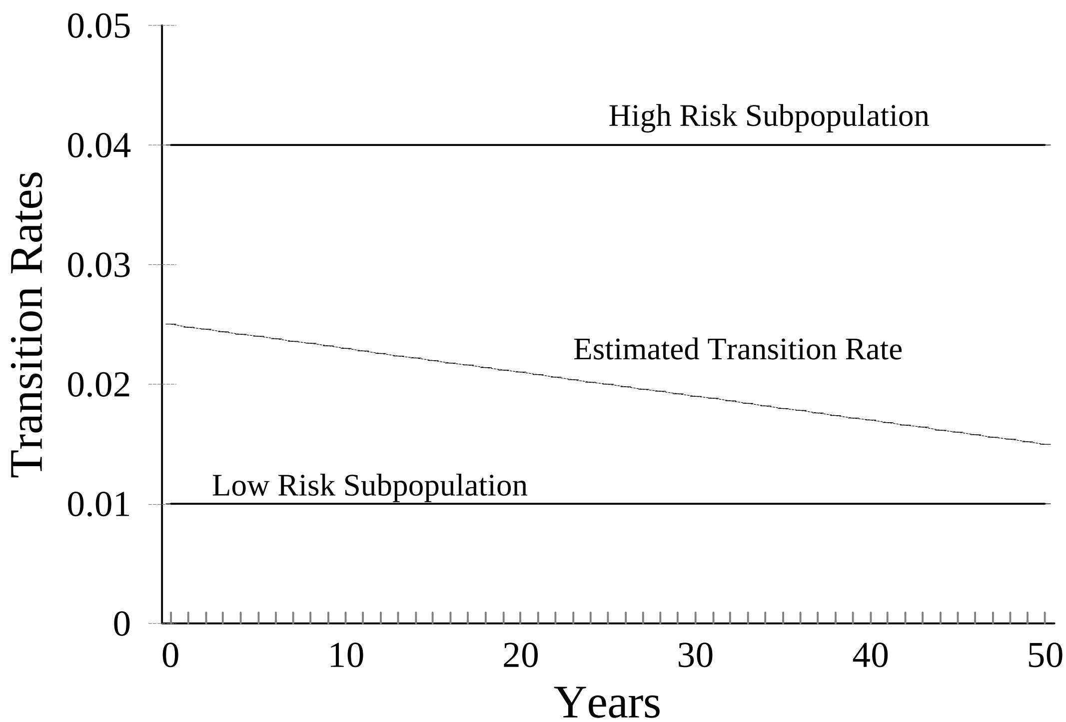
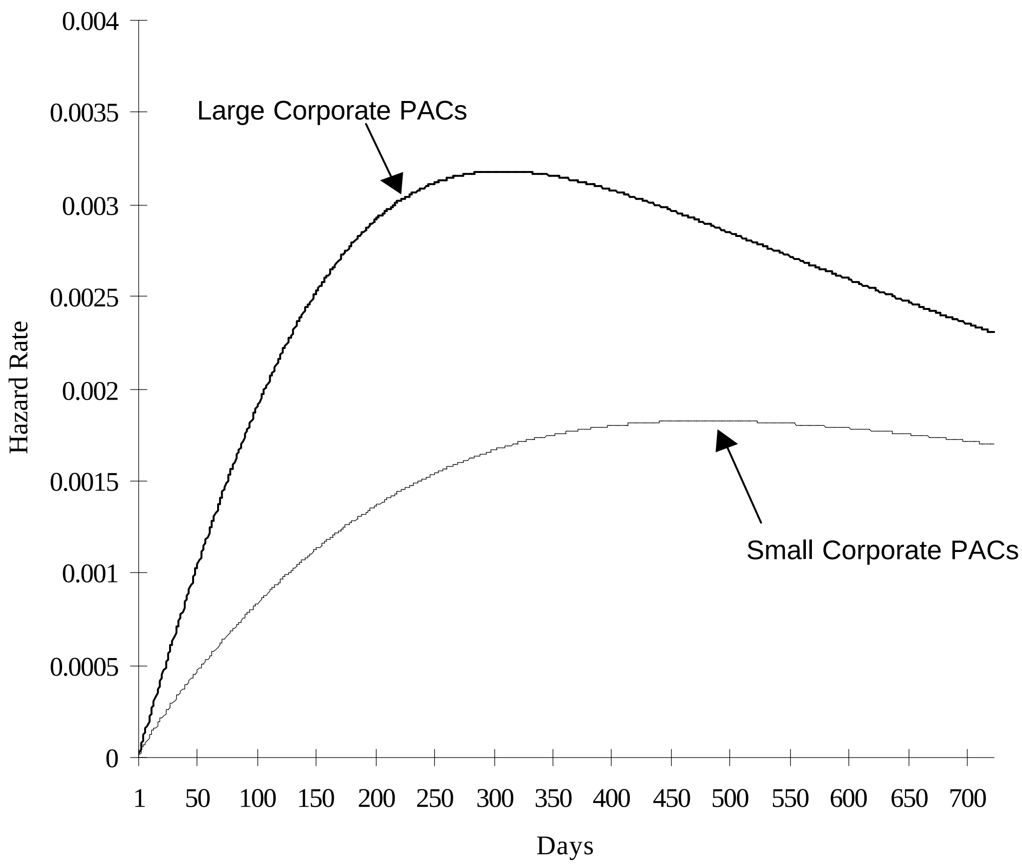


Figure 2

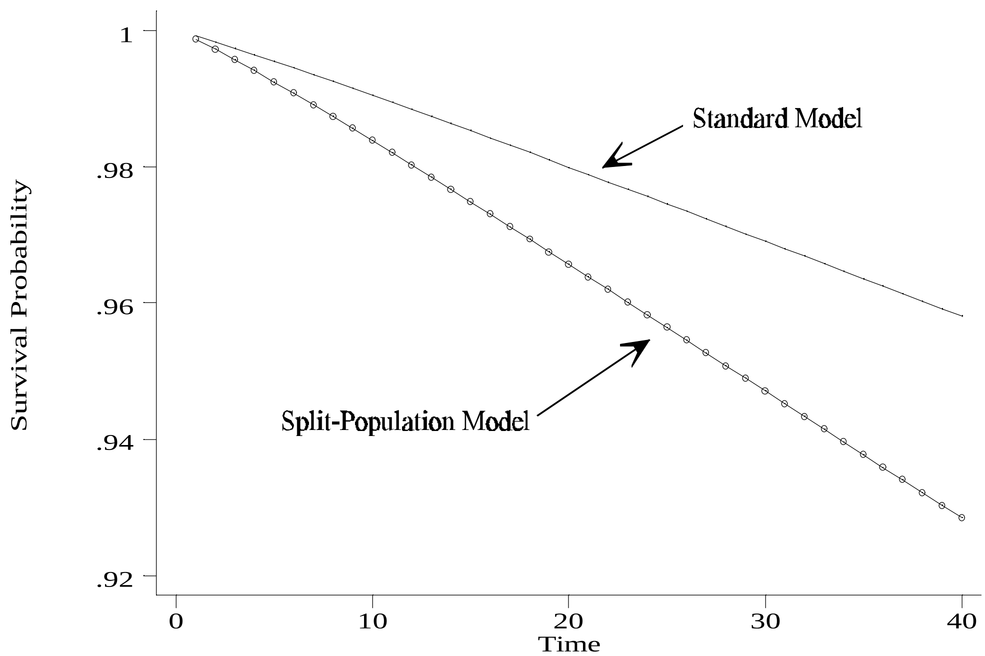
Estimated Baseline Hazards: Split-Population Models



Note: Figure plots estimated baseline hazards for small and large corporate PACs, as estimated in Table 1. See text for details.

Figure 3

Standard and Split-Population Survival Estimates



Note: Estimates are the cumulative survival probabilities for a median case, and are based on results reported in Table 3; see text for details.

Figure 4

Schematic of Approaches to Heterogeneity

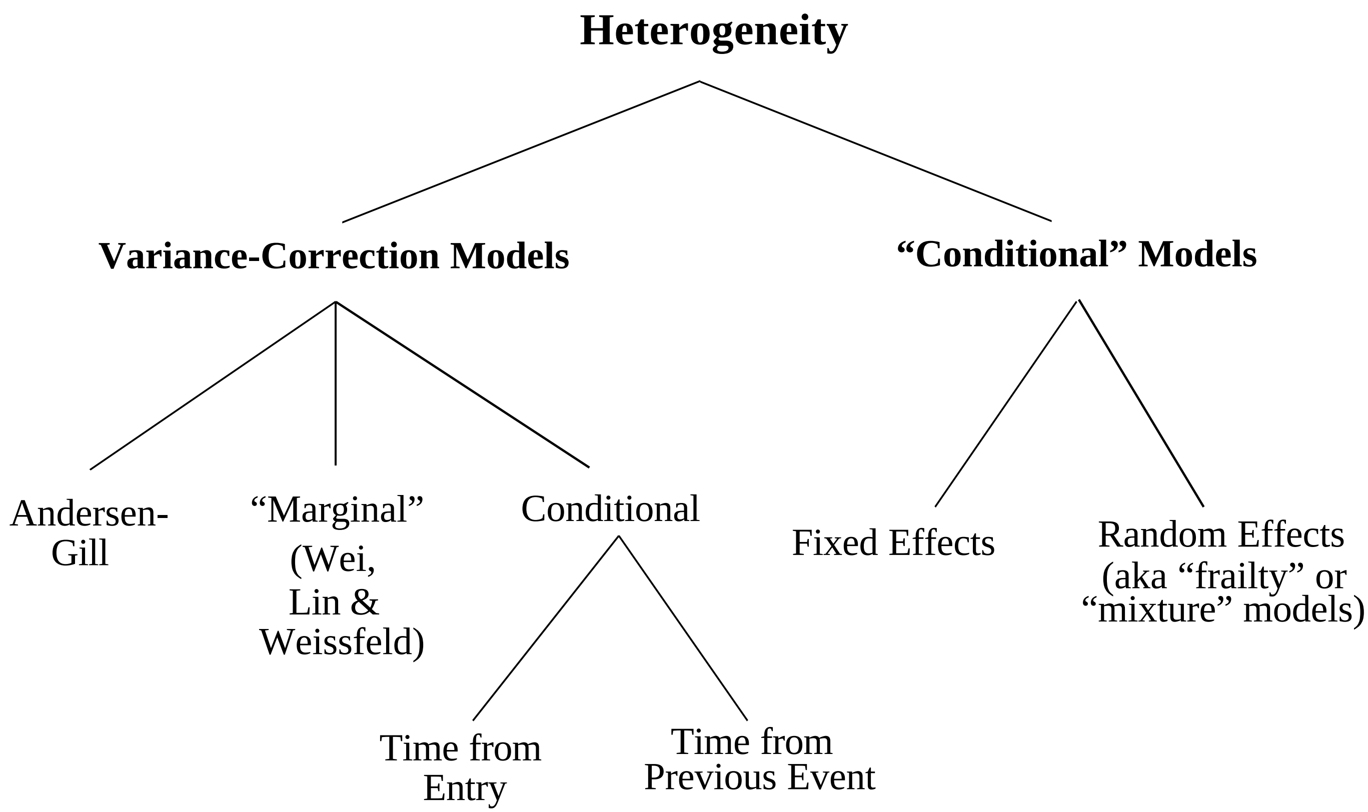
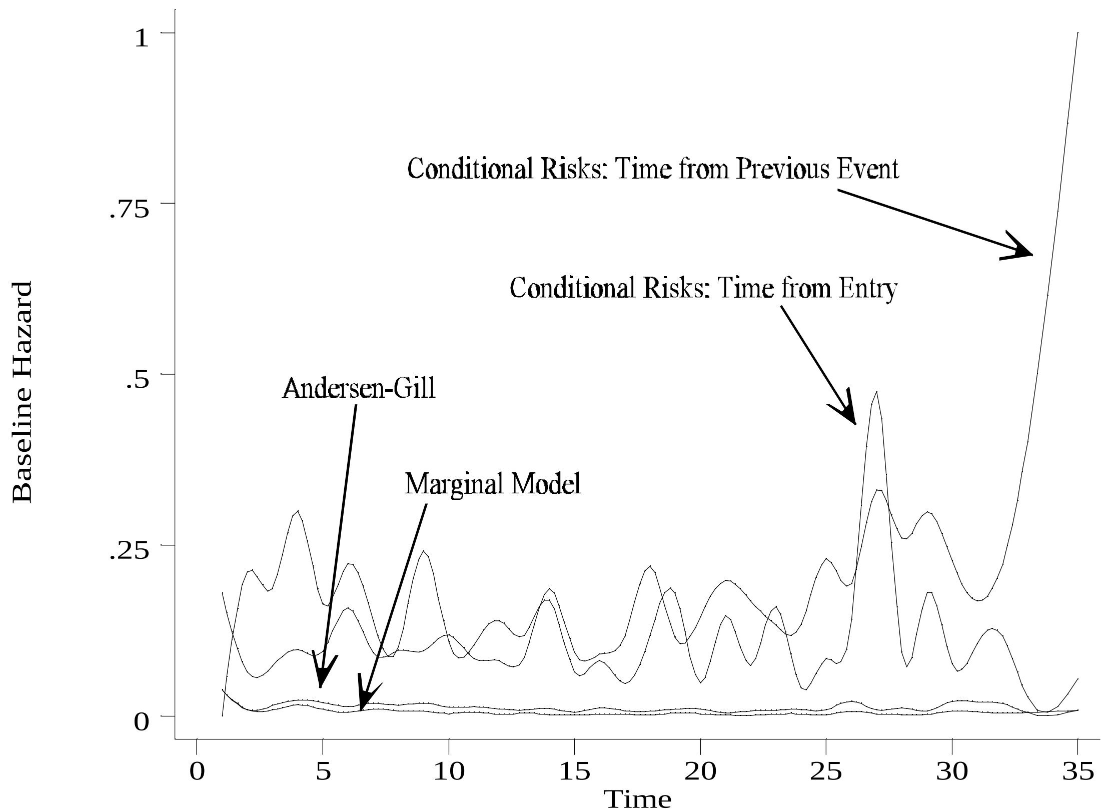


Figure 5

Baseline Hazards for Variance-Corrected Models



Note: Figures are cubic splines of the estimated baseline hazards; see text for details.