Chapter 9
Structural Interdependence and Unobserved Heterogeneity in Event History Analysis

Daniel J. Blake, Janet M. Box-Steffensmeier, and Byungwon Woo

Abstract This chapter introduces how latent variables are handled in event history analysis, a popular method used to examine both the occurrence and the timing of events. We first emphasize why event history models are popular and what kinds of research questions the model can be used to answer. We also review the major estimation issues, briefly trace the development of event history models, and highlight the differences and similarities across various types of event history models. We then consider how latent variables are handled in event history analysis and demonstrate this with an example of latent variable analysis. In the conclusion we consider possible areas for future research.

9.1 Introduction

Event history models focus on the duration of time until an event of interest occurs. An event is commonly defined as a qualitative transition from an original state to a destination state at a specific point in time. An event history is a longitudinal record of the time until an event happens (or does not happen) for each observation. Event history models have become a popular method of empirical investigation and have been widely used in many scientific disciplines.

Daniel J. Blake
Department of Political Science, Ohio State University, Columbus, USA
e-mail: blake.165@polisci.osu.edu

Janet M. Box-Steffensmeier
Department of Political Science, Ohio State University, Columbus, USA
e-mail: steffensmeier.2@polisci.osu.edu

Byungwon Woo
Department of Political Science, Ohio State University, Columbus, USA
e-mail: woo.54@polisci.osu.edu

K. van Montfort et al., *Longitudinal Research with Latent Variables*,
DOI 10.1007/978-3-642-11760-2_9,

There has been significant progress in the development of estimation techniques for event history models since the 1960s that have led to a broad range of scientific disciplines such as biostatistics, mechanical engineering, labor economics, demography, criminology, and political science to utilize event history models to study a diverse range of research questions. One of the more recent and challenging directions in the development of event history techniques has been the development of estimation approaches for multiple events. Scholars have addressed the repetition of events as well as the possibility for an individual observation to experience multiple different events (competing risks). Importantly, scholars have also sought to uncover structural independencies amongst multiple events and, in this context, have endeavored to take into account unobserved heterogeneity and the effects of latent variables.

Simultaneous equation modeling is the most prominent approach to handling structural interdependencies and unobserved heterogeneity in an event history context. One sees applications of this approach in demography, sociology, labor economics, finance, political science, and transportation engineering, particularly where scholars have modeled systems of multiple event history equations believing that multiple event history processes are interdependent, i.e. the time to one event depends on the time to another related event. Studies have examined structural interdependence between the duration of marriage and fertility timing, between the duration of breast-feeding and the duration of maternal leave, between the competing risks of getting jobs from old or new employers, and between trip and stop times in shopping activities. In short, the need for simultaneous duration models is widespread and we pay particular attention to simultaneous event history models in this chapter. However, scholars have also jointly estimated event history and non-event history models and have pursued modeling strategies such as seemingly unrelated regression (SUR) in such cases to reveal the effects of latent variables on their outcomes of interest. We address these approaches as well and replicate an existing study that employs a SUR approach to study the direction and timing of U.S. legislators' positions towards ratification of the North American Free Trade Agreement (NAFTA).

The chapter is organized as follows. In the next section, we introduce the basic elements of event history models, emphasizing why event history models are popular and what kind of research questions can be answered using event history techniques. We also briefly review several important estimation issues, briefly trace the development of event history models and highlight the differences and similarities across different types of event history models. In the third section, we consider how latent variables are handled in event history analysis. In the fourth section, we provide an example of latent variable analysis in an event history context before concluding the chapter with a discussion of possible areas for further methodological developments.

9.2 Event History Models

9.2.1 Event History Models

Event history models focus on the time until an event of interest occurs. An event is a change from one state to another, such as marriage (from single to married), divorce (from married to divorce), war (from peace to military conflict), or unemployment (from employed to unemployed). An event history is then a longitudinal record of when events of interest happened, such as a time until one's marriage or a time until one's divorce. Event history models take into account not only whether or not the event of interest occurs to an observation, but also when the event occurs and allow an investigation into timing of the occurrence of the event. The dependent variable in an event history model is the time until the event occurs. Event history models are also referred to as duration, survival, failure time, and reliability models.

Event history models have many attractive features that make them a popular choice for researchers. Compared to models that allow researchers to investigate the occurrence of an event, such as binomial logit or probit models, event history models provide opportunity to exploit the rich data of the "histories" of events in addition to the occurrence of events. As the histories provide valuable additional information, event history models help researchers better understand the causes and processes of the event of interest. For instance, a labor economist might be interested in understanding the dynamics of employment. One can see that how long it takes a job seeker to find a new job is valuable information in addition to the information about the occurrence of new employment. Event history models allow the labor economist to investigate not only what makes employment more likely but also what makes an individual more likely to find employment sooner rather than later.

Modern event history techniques can handle censored observations nicely and easily incorporate time varying covariates. One may think that ordinary least squares (OLS) regression might be able to capture factors influencing duration quite nicely, as it allows researchers to study continuous dependent variables. But event history data pose several challenges for traditional OLS regression. For one, duration data are often right skewed and the OLS approach requires an arbitrary transformation of data. A more serious problem is data truncation. Data truncation happens when researchers do not know either the exact entry time of an observation (left truncation) or the end time (right censored). Right censoring is present in almost all event history data sets as there are often observations that have not experienced an event of interest by the time data is collected. It is a problem because it results in a missing value for the dependent variable (time until event). For example, one may be interested in what causes former inmates to commit another crime and return to jail. To study this, researchers may collect data for a year after inmates are released. For those inmates that commit crimes during that year, the researchers obtain values for time until inmates re-offend. A researcher using OLS needs to treat left truncated observations as if they have the equivalent entry time with other observations and to deal with right censored observation either by dropping all the observations that have not experienced the event or by capping the history by assuming the event has occurred at the conclusion of the period of data collection. Both of these arbitrary assumptions will cause biases. A second problem for cross-sectional OLS estimation of time until events occur emerges from the potential for the values of some independent variables to change as time passes.

Event history models can handle cases of left truncation and right censoring and can accommodate time varying covariates (TVCs). Due to these advantages over more common statistical methods, event history models have garnered increasing popularity among researchers from diverse disciplines. Biostatisticians have used event history models to study the effects of medical treatments on patients' recovery time after suffering a particular disease. In clinical studies, event history models are commonly called survival models, because they are often used to study the survival of patients. In engineering, event history models have been applied to investigate times until machines or some electronic components break down. Thus, event history models in engineering are often referred to reliability or failure time models. Economists have used event history models to study durations of employment and unemployment, demographers have used them to study durations of education, and time until marriage or child bearing. Meanwhile, criminologists have used event history techniques to study the time for released inmates to commit another crime and political scientists use them to investigate such diverse topics as the timing of the dissolution of coalition governments, the breakdown of cease-fire or peace agreements between countries, and candidates' decision to run for an election.

9.2.2 Key Contributions to the Development of Event History Methods

As early as the early 1900s, life tables were used by actuaries. In the late 1950s and early 1960s, more modern methods for event history analysis were actively developed by biomedical scientists and engineers. The former developed these methods to analyze survival data gathered through clinical trials while the latter needed new statistical techniques to analyze data on the breakdown of machines and electronic components (Allison 1984, 11–12). These two research traditions effectively merged in the 1970s and, as noted earlier, event history methods have since been employed in a wide range of disciplines.

Fleming and Yin (2000) provide a summary of the most important developments in event history modeling, focusing on the work done in biostatistics. Chief amongst these is the development of the Kaplan-Meier method (Kaplan & Meier 1958) "for estimating the survival function, log-rank statistic for comparing two survival distributions (Mantel 1966)", and the Cox proportional hazard model for "quantifying the effects of covariates on the survival time" (Cox 1972). Oakes' (2001) also credits these contributions with forming the foundation of modern event history techniques, noting that "Kaplan and Meier (1958) who formalized the product-limit estimator and Cox (1972) who introduced the proportional hazards model, are primarily responsible" for the present state of art of event history models. The Kaplan-Meier estimator solved the problem of estimating a distribution function with censored data via nonparametric maximum likelihood. The comparison of two survival distributions was critical given the need to provide reliable comparisons of two populations, such as whether a medical intervention led to longer life outcomes. Cox's work has been called ingenious for his semiparametric approach that allows assessment of the influence of covariates on censored outcomes. Cox leaves the baseline hazard function unspecified and discarded the times of observed events and the number of events at those times along with an assumption that censoring is independent and uninformative. This means that the partial likelihood is based on the cases that fail at each event time given the number failing and the number of cases at risk at that time. Fleming and Yin also highlight the importance of the counting process martingale theory pioneered by Aalen (1975, p. 971) as "providing a unified framework for studying the small- and large-sample properties of survival analysis statistics." This is because it allows "one to obtain simple expressions for moments of complicated statistics and asymptotic distributions for test statistics and estimators and to examine the operating characteristics of of censored data regression method" (Fleming & Lin 2000). These important statistical developments have been instrumental to the spread of event history methods across multiple fields.

Social scientists were largely unaware of earlier developments in biostatistics and engineering. A turning point for sociology comes in the late 1970s, when Tuma (Tuma 1976) introduced "explanatory variables into continuous time Markov models, an innovation that effectively bridged the gap between the sociological approach and what had already been done in biostatistics and engineering" (Allison 1984, p. 12). In economics, early applications of event history models appeared in the late 1970s and were mostly used to explain labor force dynamics. In other social science disciplines, the adoption of event history models came later. For example, event history techniques have become increasingly popular in political science since the 1990s thanks to the work of Box-Steffensmeier who introduced event history models to the field. Software packages for survival data analysis have been widely available since the early 1980s (Allison 1984) and, currently, many common software packages support estimations of event history models.

9.2.3 Basic Elements of Event History Models

There are a wide variety of event history models, but all event history models share the same structure. There are some important technical differences between continuous and discrete time models, and between parametric and nonparametric models, but the following structure and its basic elements are shared by all. Our discussion here employs continuous time notation.

Let T be a single lifetime variable. T can be thought of as the time until an event happens and can range from 0 to a theoretical end point. Let $f(t)$ denote the probability density function of T. Then $f(t)$ denotes the probability of the event of interest occurring at any given time point t where t is an element of T. The cumulative density function of T can be obtained by integrating $f(t)$ from 1 to t.

$$F(t) = Pr(T < t) = \int_0^t f(x)\,dx \tag{9.1}$$

This is the probability of the event having occurred between time 0 and t. Then the probability of survival or the probability for an observation not experiencing the event until t can be obtained by simply subtracting $F(t)$ from 1. The probability of an individual surviving to time t is referred to as the survival function.

$$S(t) = Pr(T \geq t) = 1 - F(t) \tag{9.2}$$

The hazard rate $h(t)$ is the probability of the event happening at time t given the observation has not experienced the event until time t. In terms of the equations introduced above, the hazard rate $h(t)$ is equal to $f(t)/S(t)$, the conditional probability of an event occurring given it has not happened up until time t.

$$h(t) = \lim_{\Delta x \to +0} \frac{Pr(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} \tag{9.3}$$

The hazard is an unobserved variable, much the same way that the $Pr(Y = 1)$ is an unobserved variable in binomial logit or probit models, yet the hazard controls both the occurrence and the timing of events.

9.2.4 Different Models of Event History

There are both parametric and semiparametric event history models. The main difference between parametric and semiparametric models is that parametric models make assumptions about the structure of the baseline hazard rate once covariates are included in the model. In comparison, semiparametric models do not make such assumptions. The choice between parametric and semiparametric approaches depends on how confident researchers are of the shape of the baseline hazard, which ideally is guided by theory.

There are a wide variety of parametric models and some models are nested in other, more flexible models with more parameters. For instance, the exponential model assumes that the baseline hazard is flat across time. This means that the probability of an event occurring at time t conditional on the event not having occurred is constant over time. The exact value then depends on included covariates.

The Weibull model is more flexible than the exponential model and it allows the baseline hazard rate to be monotonically increasing, monotonically decreasing, or flat over time. This is done by inserting a linear function of t into the right hand side of the equation. When the coefficient of the t term is 0, the Weibull model becomes the exponential model. Thus, the Weibull model is nested in the exponential model.

Since the Weibull model is flexible, it is a popular choice in social science applications. Yet, in some settings, the monotonicity assumption may not be appropriate.

When one suspects that the monotonicity assumption is not defendable, the log-logistic and the log-normal models can be used. These models allow hazard rates to first increase and then decrease as t passes. Neither of these models have the proportional hazards property. The generalized gamma model can be useful to adjudicate among different parametric models as several parametric models are nested within the generalized gamma model. The exponential, the Weibull, the log-normal, and the gamma models are all special forms of the generalized gamma model. When one has no a priori theoretical justification about how the baseline hazard rate varies across time, the generalized gamma model is more likely to be useful. If the fit is correct, parametric models generally have smaller standard errors than their semi-parametric counterparts. On the other hand, as parametric models require a priori assumptions about the shape of the baseline hazard, when assumptions are not correct, the estimation will be biased.

Recently, the Cox (1972, 1975) semiparametric model has become the most commonly used in social science applications (Therneau & Grambsch 2001, Singer & Willett 1993, Box-Steffensmeier & Jones 2004). The Cox semiparametric model's primary advantage is that it relaxes the assumption that the time until an event occurs follows a specific distribution. Larsen and Vaupel (1993) point out that "in the analysis of duration data, if the functional form of the hazard has the wrong shape, even the best-fitting model may not fit the data well enough to be useful" (p. 96).

A key concept for understanding the Cox model is the hazard rate. Recall, that the hazard rate can be thought of as the probability that an event will occur for a particular observation at a particular time, or the rate at which an event occurs for an observation at time t given that the observation has survived through time $t-1$. In the Cox model, the hazard rate for the ith individual is

$$h_i(t) = h_0(t)\exp(\beta'\mathbf{x}), \tag{9.4}$$

where $h_0(t)$ is the baseline hazard function, and $\beta'\mathbf{x}$ are the covariates and regression parameters. A Cox model does not report an intercept as it is absorbed into the baseline hazard function. The ratio of two hazards (or hazard ratio) can be written as,

$$\frac{h_i(t)}{h_0(t)} = \exp(\beta'(\mathbf{x}_i - \mathbf{x}_j)), \tag{9.5}$$

which demonstrates that this ratio is a fixed proportion across time. Box-Steffensmeier and Jones (2004) point out that when the proportional hazards assumption holds in the Cox model, the particular form of the baseline hazard rate, $h_0(t)$ is assumed to be unknown and is left unparameterized. More accurately the duration times are parameterized in terms of a set of covariates, but the particular distributional form of the duration times is not parameterized, hence the term "semi-parametric".

Proportional hazards is the major assumption of the Cox model, as well as many parametric models. This assumption is tested with both the global model test defined by Therneau and Grambusch (1994) and Harrell's rho for individual covariates (Box-Steffensmeier & Jones 2004, 135). If the assumption is found to be violated, the offending covariate(s) is interacted with time and the model is re-estimated. Like all models, there is a series of general diagnostics for the Cox model. These include assessments of model fit, functional form, and influence (Therneau, Grambsch, & Fleming 1990, Grambsch, Therneau, & Fleming 1995, Grambsch & Therneau 1994).

Parameters in the Cox model are estimated using a *partial likelihood* approach. The partial likelihood method is based on the assumption that the intervals between successive duration times (or failure times) contributes no information regarding the relationship between the covariates and the hazard rate (Collett 2003), which comports to the arbitrary form assumed for the baseline hazard. Because the Cox model only uses "part" of the available data ($h_0(t)$ is not estimated), the likelihood function for the Cox model is a "partial" likelihood function. In contrast, consider the more typically encountered likelihood function which gives hypothetical population value that maximizes the likelihood of the observed sample using all of the data. That is, the maximum likelihood estimate is the value that is the most likely to generate the sample that is observed.

To derive the partial likelihood function for a data set of size n with k distinct failure times, the data are first sorted by the ordered failure time, such that $t_1 < t_2 < \ldots < t_k$, where t_i denotes the failure time for the ith individual. For now, we assume that there are no "tied" events: each uncensored case experiences an event at a unique time. For censored cases, we define δ_i, see 9.7, to be 0 if the case is right-censored, and 1 if the case is uncensored, that is, the event has been experienced. Finally, the ordered event times are modeled as a function of covariates, $\mathbf{x}$.

The partial likelihood function is derived by taking the product of the conditional probability of a failure at time t_i, given the number of cases that are at risk of failing at time t_i. More formally, if we define $R(t_i)$ to denote the number of cases that are at risk of experiencing an event at time t_i, that is, the "risk set," then the probability that the jth case will fail at time T_i is given by

$$\Pr(t_j = T_i \mid R(t_i)) = \frac{e^{\beta' \mathbf{x}_i}}{\sum_{j \in R(t_i)} e^{\beta' \mathbf{x}_j}}, \tag{9.6}$$

where the summation operator in the denominator is summing over all individuals in the risk set. Taking the product of the conditional probabilities in (9.6) yields the partial likelihood function,

$$\mathscr{L}_p = \prod_{i=1}^{K} \left[\frac{e^{\beta' \mathbf{x}_i}}{\sum_{j \in R(t_i)} e^{\beta' \mathbf{x}_j}} \right]^{\delta_i}, \tag{9.7}$$

with corresponding log-likelihood function,

$$\log L_p = \sum_{i=1}^{K} \delta_i \left[\beta' \mathbf{x}_i - \log \sum_{j \in R(t_i)} e^{\beta' \mathbf{x}_j} \right]. \tag{9.8}$$

By maximizing the log-likelihood in (9.8), estimates of the β are obtained. As (Collett 2003) notes, the likelihood function in (9.7) is not a true likelihood. This is because the actual survival times of censored and uncensored cases are not directly incorporated into the likelihood. Nevertheless, Cox (1972, 1975) famously demonstrated that maximum partial likelihood estimation produces parameter estimates that have the same properties as maximum likelihood estimates - asymptotically normal, asymptotically efficient, consistent, and invariant (see also Collett 2003).

The logic underlying the partial likelihood method is seen by considering the data presented in Table 9.1 (this part of the presentation is directly adapted from (Collett 2003)). We reproduce this here because of the clarity of Collett's example. The survival times for nine cases are provided. Of these nine cases, six of them experience an event, i.e., they "fail", and three of them are right-censored. The failure times can be ordered such that $t_1 < t_2 < \ldots < t_6$. Note that the censored cases do not contribute a failure time. Each of the nine cases are at risk of experiencing an event up to the first failure time, t_1. After the first failure in the data set, the risk set decreases in size by 1; thus, the risk set up to the second failure time, t_2, includes all cases except case 7. By the fourth failure time in the data, t_4, the risk set includes only cases 1, 6, and 8; cases 2 and 9 are right-censored before the fourth failure time is observed and do not contribute any information to this part of the likelihood function. By the last failure time, only case 6 remains in the risk set. Using the notation from Collett (2003, 64), let $\psi = \exp(\beta' \mathbf{x}_i)$. Then the partial likelihood function for these data would be equivalent to

$$\begin{aligned}
\mathscr{L}_p = \; & \frac{\psi(7)}{\psi(1)+\psi(2)+\psi(3)+\psi(4)+\psi(5)+\psi(6)+\psi(7)+\psi(8)+\psi(9)} \times \\
& \frac{\psi(4)}{\psi(1)+\psi(2)+\psi(3)+\psi(4)+\psi(5)+\psi(6)+\psi(8)+\psi(9)} \times \\
& \frac{\psi(5)}{\psi(1)+\psi(2)+\psi(3)+\psi 5+\psi(6)+\psi(8)+\psi(9)} \times \\
& \frac{\psi(3)}{\psi(1)+\psi(3)+\psi(6)+\psi(8)} \times \\
& \frac{\psi(1)}{\psi(1)+\psi(6)} \times \\
& \frac{\psi(6)}{\psi(6)}.
\end{aligned}$$

Again we see that the partial likelihood function is based on ordered duration times and censored observations contribute information to the "risk set" but contribute

Table 9.1 Data Sorted by Ordered Failure Time

Case Number	Duration Time	Censored Case
7	7	No
4	15	No
5	21	No
2	28	Yes
9	30	Yes
3	36	No
8	45	Yes
1	46	No
6	51	No

no information regarding failure times. In terms of the likelihood function in (9.7), censored observations contribute information to the denominator, but not to the numerator.

"Ties," or coterminous event occurrences, cannot be accounted for in the partial likelihood function, as presented in (9.7). This is true for any continuous time model. However, the literature has adapted a number of approximations to take this into account. For example, numerically computing "what if" this tied event occurred first, then the computing the same "what if" for the next tied event and so on. The Efron method is a popular choice for handling ties.[1]

In sum, event history models allow scholars to more fully capture the process surrounding the occurrence (or nonoccurrence) of events. We can investigate whether covariates speed up or slow down the timing of events and gain a more complete understanding of the process with event history models.

9.3 Statistical Models for System of Equations

In general, simultaneous equation models are used when there is a system of relationships, such as a two-way flow of influence. For simplicity, consider a variable A that affects another variable B and that is also affected by variable B. In this case, we need to consider a two equation setup where there is one equation for each interdependent or endogenous variable. When estimating the parameters for simultaneous equation models, information from both (or all if there are more than two endogenous variables) equations have to be taken into account. If this is not done, biased and inconsistent estimators are the result.

A major hurdle for estimation of simultaneous equations is the identification problem. In short, the identification problem arises because the same set of data may be compatible with different models. The identification problem needs to be

[1] See Golub and Collett (2002) for further discussion of this issue of ties.

addressed before considering estimation strategies. Identification occurs through the introduction of a priori information into the analysis. There are a variety of techniques for estimating simultaneous equations and it is a vibrant and long-standing area of econometric research, as is shown in this volume and in van Montfort, Oud, and Satorra (2004).

Simultaneous Equation (SEQ) models are used when the equations for each part of the system are interdependent (also referred to as a substantive approach). Arguably, such interdependence is ubiquitous in the world around us and the questions studied by social scientists. In this case, it is more than just the disturbances that are related. For a complete system of equations, the number of equations needs to equal the number of endogenous variables. Joint estimation of all the equations in the system provides fully efficient approaches.

The Seemingly Unrelated Regression (SUR) model is an alternative approach to SEQ estimation of interdependent processes. The SUR model has a series of equations that are linked through correlated error terms (also referred to as a nuisance approach) and generalized least-squares (GLS) estimation is used to gain efficiency. The higher the correlation of the disturbances, the greater the efficiency gain in using GLS (Zelner 1962, Dwivedi & Srivastava 1978). SUR estimation is "simply the application of generalized least-squares estimation to a group of seemingly unrelated equations. The equations are related through the nonzero covariances associated with error terms across different equations at a given point in time" (Pindyck & Rubinfeld 1991, 326). Both autocorrelation and heteroscedasticity can be accommodated in the SUR model.

Both the SUR and SEQ approaches are central to the estimation of the effects of latent variables in an event history context. However, the choice of approach depends on the type of interdependence between processes the researcher assumes is present. We discuss the choice between different approaches in more depth and illustrate how these techniques have been used in the extant literature to ascertain the presence and effect of latent variables in the next section.

9.4 System of Equations, Interdependent Processes and Latent Variable Analysis

There are numerous occasions where we want to study the duration of an event within a framework of a system of equations using either SEQ or SUR approaches. First, we may wish to model multiple endogenous event history processes simultaneously. This can occur in two different ways depending on the relationship between the events of interest. In many cases, scholars will be interested in jointly modeling the duration processes for multiple events where the events of interest are not mutually exclusive. Lillard (1993) conceives of this as modeling "multiple clocks," which refers to one process depending on the duration of a related process. For example, one may be interested in modeling the time until a woman completes her education and the time until a woman becomes pregnant. The time until a woman completes her education may be affected by whether she is pregnant or not and the timing of a women's becoming pregnant may be affected by whether she completes her education or not. There are plenty of research questions that features multiple endogenous event histories: marital duration and the time to marital conceptions; duration of breast-feeding and duration of maternal leave; and duration of women's education and the time to entry into a first union.

In other contexts, events may appear as competing risks for a common duration process. Compared to multiple event history processes, competing risks models have a single duration process that can end with multiple events, whereby the occurrence of one event necessarily rules out the occurrence of another event. For example, one could model the duration of military conflict as a competing risks process: military conflict can end with an invading country's victory or a defending country's victory. As two states cannot win simultaneously on the battlefield, when an event (winning by a country), occurs, it is not longer at risk of experiencing the other event. In the simultaneous equations context, the competing risks can be related to each other. For instance, an invader's decision to continue fighting for a victory may be dependent on a defending country's decision to continue fighting and vice versa. In this case, the two hazards of the two competing risks need to be jointly estimated. Rosholm and Svarer (2001) estimate unemployment durations with two competing risks: the risk of being recalled by the previous employer and the risk of being hired by a new employer. As they theoretically expect that the hazard of getting a new job is dependent on the hazard of being recalled by the previous employer (the hazard for recall should reduce the hazard of new jobs as those who see higher probability of being recalled will be less active in pursuing new jobs), they put the hazard for recall in estimating the hazard for new jobs when constructing a system of equations and estimate structural dependency between the two hazards. The simultaneous competing risks models are useful in many situations: duration of economic sanctions where the duration can end either with target's capitulation or sender's lifting economic sanctions; duration of hospitalization where the duration can end with different events.

When jointly estimating multiple event histories, the equations in the system are all structured as some form of duration model. However, we may also wish to model duration processes jointly with other non-duration processes. In many cases, the non-duration model we wish to model attempts to estimate some important aspect of the event itself. For instance, we may wish to simultaneously model the time until a government calls an election and the result of that election, whereby timing is modeled using an event history technique and the result, measured as vote share for the incumbent government, is estimated using ordinary least squares regression (see Fukumoto 2009).

In sum, SUR and SEQ approaches may be usefully employed in event history settings when scholars are interested in estimating a system of interrelated event history models, a system of event history models with mutually exclusive outcomes, and a system of models containing both event history and non-duration equations.

When deciding which specific modeling strategy to pursue in each of these three cases, researchers need to be clear about their assumptions regarding the nature of interdependence between their multiple processes of interest. If scholars believe the processes are independent purely through their stochastic components, than a SUR approach may be appropriate as a SUR approach assumes that outcomes are interdependent because the errors of both processes share a single joint probability distribution (Hays & Kachi 2009). With respect to latent variables, since a SUR approach focuses on correlation in the errors terms, it allows scholars to identify the presence or absence of unobserved factors acting upon the dependent variables of interest. Fred Boehmke's (2006) study of the timing and position of U.S. legislators' towards the ratification of the North Atlantic Free Trade Agreement (NAFTA), which we replicate below, is an excellent example of a SUR approach. Boehmke argues that unobserved bargaining dynamics and competing pressures on legislators jointly influenced their positioning and timing on NAFTA, causing the two processes to be positively related and he finds evidence that Democratic legislators who declared their positions later in time, also tended to come out in favor of NAFTA. In another study, Fukumoto (2009) uses copula techniques to model dependence in latent variables between event history models and models of the event themselves. One advantage of copula approach is that asymmetric interdependence can be captured and modeled. Asymmetric interdependence occurs when one actor/process is more dependent on a second actor/process than the second actor/process is dependent on the first. For example, unobserved heterogeneity in the duration and event models may be such that latent variables in the duration model affect the event, but not vice versa. Both Fukumoto and Boehmke employ SUR approaches to model interdependent processes where only one of those processes is an event history process. However, in another study employing copula functions, Quiroz Flores (2008) estimates two event history models - the tenure in office of chief executives and the tenure of their foreign ministers - and finds that the tenure of individuals in both offices are closely correlated.

The main alternative to a SUR approach is to generate simultaneous equation models (SEQ) of interdependent processes of interest (Hays & Kachi 2009). The SEQ is preferable when endogeneity extends beyond stochastic components and one wishes to explicitly model the interdependence among outcomes of interest. In a system of simultaneous equations, endogenous variables appear on the right hand side of the equations which has implications for our analysis of the influence of unobserved, or unobservable, variables. This is because the variances an covariances among errors, in the reduced form of the structural equations, "need to be consistent with the structural relationship among endogenous variables" (Hays & Kachi 2009, p. 4).

The SEQ approach is by far the most popular approach in the extant literature for scholars interested in modeling multiple interdependent duration processes. Many researchers have chosen to build off the model developed in Lillard (1993) which has been quickly established as a classic article on combining simultaneous equations and duration models. Lillard (1993) presents a comprehensive model of the dynamics of marriage duration and marital fertility, i.e., the timing of marital conceptions taking into account a number of time-varying covariates, a set of exogenous covariates, and a set of endogenous covariates. He proposes that the hazard of conception if influenced by the hazard of marriage dissolution along with other variables and conversely, the hazard of marriage dissolution is influenced by the duration and outcome of marital fertility. As is common in the demography literature, he models the baseline hazard with a Gompertz distribution, then builds a system of equations, and obtains the reduced form equation.[2] As there are variables that influence one hazard but not the other, he can obtain identification of the structural dependence parameter. Lillard addresses the across-duration interdependence by allowing the baseline hazards to take flexible shapes through the use of splines.

The errors in Lillard's model represent residual, unobserved heterogeneity and their distribution is conditioned on the relationship among the endogenous variables. Joint normality of the error terms in the two separate hazard equations of fertility and marriage, is assumed[3],

$$\begin{pmatrix} \varepsilon \\ \eta \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon\eta} \\ \sigma_{\varepsilon\eta} & \sigma_\eta^2 \end{pmatrix} \right) \tag{9.9}$$

where ε is the error term in the hazard equation for dissolution of marriage, η is the error term in the hazard equation for conception (Lillard 1993). No assumption is made regarding the correlation or lack thereof between ε and η. However, the structural relationship between endogenous variables introduces some correlation in the residuals, which, in the reduced form of the simultaneous equation, are distributed as follows:

$$\begin{pmatrix} \varepsilon \\ \eta + \lambda\varepsilon \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon\eta} + \lambda\sigma_\varepsilon^2 \\ \sigma_{\varepsilon\eta} + \lambda\sigma_\varepsilon^2 & \sigma_\eta^2 + 2\lambda\sigma_{\varepsilon\eta} + \lambda^2\sigma_\varepsilon^2 \end{pmatrix} \right) \tag{9.10}$$

where λ is the coefficient for the endogenous hazard of marital disruption in the hazard model for conception. Thus, it is through the inclusion of λ that the distribution of errors is consistent with the structure of endogeneity between marriage dissolution and conception. It is here that the distinction between SUR and SEQ is perhaps most important for scholars who wish to investigate the presence of latent variables through examining the relationship between the residuals of multiple processes of interest. A SUR approach assumes no endogeneity in independent variables and therefore λ is never estimated.

Scholars who have modeled simultaneous duration processes, but are interested in the impact of latent variables have frequently followed Lillard's modeling approach and then investigated whether or not there is correlation in the heterogeneity terms, or errors. For example, Baizán, Aassve, and Billari (2004), interested in the time until the formation of cohabitation or marital relationships and the time until a couple's first child is born, find a positive correlation between the errors in the model of union formation and the model of childbirth. Other studies have looked at

---

[2] Olshanksy and Carnes (1993) discusses the appropriateness of the Gompertz distribution for demography based on its U-shape.

[3] Recall that the SUR approach also assumes the error terms of multiple models follow a joint distribution.

the correlation between heterogeneity components in models of duration of education and time to union formation (Billari & Philipov 2004, Coppola 2004), time to migration and to union dissolution (Boyle, Kulu, Cooke, Gayle, & Mulder 2008), and time between children and child mortality (Maitra & Pal 2007). An important attribute of the Lillard approach is that it allows scholars to both model endogenous processes and investigate the presence of correlated but unobserved heterogeneity. This makes his approach very suitable if researchers perceive their processes of interest as being "characterized by 1) mutual influence - that is events in one process trigger events in the other process - and 2) common time-constant influencing factors - which are usually not observed especially in retrospective surveys and which represent sources of potential endogeneity" (Billari & Philipov 2004, p. 92).

Many scholars adopting a SEQ approach display a theoretical interest in the endogeneity of the multiple processes, the presence of unobserved heterogeneity and the possible correlation of this heterogeneity across models. However, some scholars simply wish to statistically account for endogenous right hand side variables and the potential effects of unobserved factors. An instrumental variables approach is often pursued in such cases. For example, in their study of the duration of breast-feeding, Adair, Popkin and Guilkey (1993) argue that independent variables such as child health and the use of oral contraception influence how long mothers breast-feed their newborn infants, but are aware that the duration of breast feeding in turn shapes the health of the child and the decision to use contraception. Their methodological strategy is to generate predicted values of the potentially endogenous covariates by estimating separate OLS and logistic regression models for each endogenous variables using a battery of household factors, only some of which they include in their model of decisions regarding breast-feeding. These predicted values are then used to estimate the main discrete time logit hazard model of breast-feeding, with bootstrapped standard errors to overcome the problem of conditional standard errors. Addison and Portugal (1989) take a similar approach in their study of post-unemployment wage displacement, however, their main dependent variable of interest is not time to failure, but their potentially endogenous right hand side variable, duration of unemployment, is produced by a survival process. Thus, they generate predicted failure times for unemployment, which they then use in an OLS model of wage displacement. By enabling researchers to identify and statistically control for endogeneity, an instrumental variables approach allows them to account for potentially unobserved factors that affect both the value of endogenous right hand side variables and the dependent variable of interest.

Rosholm and Svarer (2001) also specify and estimate a simultaneous equations model for hazards. Yet, they do not examine structural dependency of two district temporal processes; they estimate the two interdependent hazards of unemployment duration. They consider unemployment ending with two exits (recall from the previous employer or a new job) and model the structurally dependent competing risks between the two processes. They find that the recall hazard affects the new job hazard negatively when taking the structural dependency into account. They suggest that the structurally dependent competing risks model is a fruitful alternative to the standard competing risks model.

There is exciting work in the field of political methodology on the topic of simultaneous duration models. In addition to the study of NAFTA positioning by Boehmke (2006) mentioned above, Boehmke, Morey, and Shannon (2006) start with Gumbel's bivariate exponential distribution to model non-random selection when the outcome of interest is duration. They apply this approach to the study of the effect of leaders' decisions to go to war on their subsequent post-crisis tenure. One of the shortcomings of the Boehmke (2006) and Boehmke et al. (2006) modeling strategy is that it does not allow researchers to identify the direction of influence between models when interdependence is asymmetric (Hays & Kachi 2009). Knowing the direction and strength of that asymmetry is often of theoretical interest.

Hays and Kachi (2009) add structure to the empirical models to estimate precise causal effects of one process on the other and vice versa. They present a simultaneous equations model for multiple interdependent duration processes using the Weibull distribution and derive its full information maximum likelihood estimator (FIML). The FIML estimator is shown to be efficient compared to a two stage least squares approach in the Monte Carlos. In their substantive application, they examine the interdependence between the duration of coalitional government formation and the duration of governmental survival. In their simultaneous equations setup, they create a system of structural equations and present the corresponding reduced form that is used to derive the likelihood function. With the estimation result of the simultaneous model, they conclude that government survival causes bargaining duration rather than the reverse, and thus that the positive covariance between the two durations is attributable to strategic bargaining. When parties expect a longer government, they bargain harder, which results in a longer negotiation duration.

Importantly, Hays and Kachi allow across-unit interdependence (in addition to across-duration dependence). A major contribution of their work is that they introduce a general approach that allows inclusion of both across-duration and across-unit dependences in one model. For example, there can be multiple interdependent duration processes, and the observational units within each duration process can be interdependent as well. They also compare the duration seemingly unrelated regression (SUR) models, such as Boehmke (2006), Boehmke et al. (2006), and Quiroz Flores (2008) and the simultaneous equations (SEQ) framework. Thus, they bring together various strands of work.

9.5 Duration and Discrete Choice: Timing and Direction of Position Taking by Legislators

We illustrate the duration and discrete choice model estimation and interpretation by replicating Boehmke (2006), who derives a seemingly unrelated discrete-choice duration estimator (SUDCD). His work is used to evaluate the presence of unobserved processes in a duration and a discrete choice context that are *not* independent. This is an excellent example of the SUR approach and a good illustration of why latent variable estimation may be important and how it can be incorporated in an event history setting. The empirical motivation for his model comes from an already existing study that employs event history techniques. This study is an article by Box-Steffensmeier, Arnold, and Zorn (1997) which explains the timing and direction of positions taken by members of the United States Congress on the ratification of the North American Free Trade Agreement (NAFTA) using two separate models: a Cox model to estimate timing and a probit model to estimate direction (for or against NAFTA). Boehmke argues that two unobservable, but strategic, processes that link together position choice and the timing of position announcement are present. First, if legislators face competing pressures regarding which position to take, they may delay selecting a position to see if their vote will have a significant effect on the final outcome before deciding which pressure to give in to. Second, legislators may delay taking a position if they are indifferent towards the outcome of the vote, but hope to induce side-payments from other representatives and actors with more at stake. Such legislators refrain from declaring a position early in case the vote should appear to be very close, thus increasing the importance of their vote and the amount of side-payments they can extract from both sides. Ultimately, these legislators will take a position based on which side offered them the best deal. These two processes are both unobservable and affect both the timing and content of a legislator's position on NAFTA.

To better evaluate these processes, Boehmke derives an estimator that follows a bivariate distribution which allows for nonzero correlation between the duration and discrete outcome equations. The Cox model cannot be used because it does not have a parametric assumption about the distribution of errors. So, two estimators are constructed using two different parameterizations of the baseline hazard that are commonly used in event history models: the Weibull and log-normal distributions[4].

9.5.1 Boehmke's Derivation of the Weibull SUDCD Estimator

The structure of the likelihood function is of the form (D_i, V_i) where D_i is the timing of a position for an individual and V_i is the position on whether to support or reject NAFTA. The duration equation is the same for individuals who reject or support NAFTA and thus the critical difference is between individuals' positions on NAFTA. The likelihood can be written using the marginal density of the duration and conditional probabilities of support for NAFTA.

[4] We address only the Weibull version in detail as it is the most common parametric model. Please note that while the Weibull's companion discrete choice model's errors follow a bivariate exponential distribution, the companion for the log-normal is a standard probit model.

$$\begin{aligned} Pr(\mathbf{D},\mathbf{V}) &= \prod_{i=1}^{n} P(D_i = d_i, V_i = v_i) \\ &= \prod_{i=1}^{n} P(D_i = d_i) \times P(V_i = 0 | D_i = d_i)^{1-v_i} \times P(V_i = 1 | D_i = d_i)^{v_i} \end{aligned} \quad (9.11)$$

For the estimator, the joint density and conditional probabilities are calculated using the bivariate exponential distribution, where the cumulative and probability density functions take the form:

$$F_{exp}(x,y) = (1 - \mathrm{e}^{-x})(1 - \mathrm{e}^{-y})(1 + \alpha \mathrm{e}^{-x-y}), \quad (9.12)$$

$$f_{exp}(x,y) = \mathrm{e}^{-x-y}[1 + \alpha(2\mathrm{e}^{-x} - 1)(2\mathrm{e}^{-y} - 1)] \quad (9.13)$$

It is worth noting that the correlation between x and y is given by $\rho = \alpha/4$ and as α is bound between -1 and 1, the value of the correlation parameter, ρ, is restricted to $-0.25 \leq \rho \leq 0.25$.

The first part of the estimator is a standard exponential duration equation that is adapted to allow for Weibull duration dependence. The exponential distribution is given by: $d_i = \exp(\mathbf{x}_i\beta)\varepsilon_i$ where $\mathbf{x}_i$ is a vector of independent variables and ε_i has an exponential distribution. The marginal density of observing a duration d_i is given by:

$$f_w(d_i|\lambda_i) = p\lambda_{2i}^{p} d_i^{p-1} \exp[-\lambda_{2i} d_i^{p}], \quad (9.14)$$

where $\lambda_{2i} = \exp(-\mathbf{x}_i\beta)$ and p is a shape parameter that causes a Weibull distributed variable, u_i, to follow an exponential distribution.

The second part of the estimator is the discrete choice equation, developed to allow the errors to follow an exponential distribution. As the exponential distribution is not defined for negative numbers, the outcome is modeled:

$$V_i = \begin{cases} 1, \text{ if } \exp(\mathbf{w}_i\gamma)\eta_i > 1 \\ 0, \text{ otherwise,} \end{cases} \quad (9.15)$$

where η_i follows an exponential distribution. The marginal probability that $V_i = 0$ can is given by:

$$\begin{aligned} P(\exp(\mathbf{w}_i\gamma)\eta_i \leq 1) &= p(\eta_i \leq \exp(\mathbf{w}_i\gamma)) \\ &= 1 - \exp(-\lambda_{1i}), \end{aligned} \quad (9.16)$$

where $\lambda_{2i} = \exp(-\mathbf{w}_i\gamma)$. The likelihood function for SUDCD can now be written out in full, combining the two parts, as follows:

$$L(\beta,\gamma,p,\alpha|\mathbf{X},\mathbf{W},\mathbf{D},\mathbf{V}) = \prod_{i=1}^{n} p\lambda_{2i}^{p} d_i^{p-1} \exp[-\lambda_{2i} d_i^{p}](1-\pi_i^1)^{1-v_i}(\pi_i^1)^{v_i}, \quad (9.17)$$

$$\ln L(\beta,\gamma,p,\alpha|\mathbf{X},\mathbf{W},\mathbf{D},\mathbf{V}) = \sum_{i=1}^{n} (\ln(p) + p\ln(\lambda_{2i}) + (p-1)\ln(d_i) - (\lambda_{2i} d_i)^{p}) + [(1-v_i)\ln(1-\pi_i^1) + v_i \ln(\pi_i^1)], \quad (9.18)$$

where $\pi_i^1 = \exp(-\lambda_{1i})\{1+\alpha[2\exp(-(\lambda_{2i}d_i)^p)-1][\exp(-\lambda_{1i})-1]\}$ is the conditional probability that V_i is one.

With respect to right-censored observations, "their contribution to the overall likelihood is the joint probability of surviving until right censoring occurs and the probability of the observed discrete-choice outcome" (Boehmke 2006, p. 7), calculated as follows:

$$Pr(D_i \geq d_i^c, V_i = 1) = 1 - F_{\exp}(\lambda_{1i}) - F_{\exp}((\lambda_{2i} d_i^c)^p) + F_{\exp}(\lambda_{1i}, (\lambda_{2i} d_i^c)^p), \quad (9.19)$$

$$Pr(D_i \geq d_i^c, V_i = 0) = F_{\exp}(\lambda_{1i}) - F_{\exp}(\lambda_{1i}, (\lambda_{2i} d_i^c)^p), \quad (9.20)$$

where d_i^c is the censoring point.

9.5.2 Practical Application: Position Taking on NAFTA

Box-Steffensmeier, Arnold, and Zorn (1997) identify a range of covariates that potentially influenced the position members of Congress took on NAFTA, when they chose to publicly announce that position, or both. These independent variables, and the expected direction of their influence on timing and direction, are summarized in Table 9.2, adapted from Box-Steffensmeier, Arnold, and Zorn[5] (1997). The reader is directed to this article and Boehmke (2006) for further discussion of the theoretical reasoning underlying the expected effects of these *observable* factors.

To investigate the effect of *unobserved*, but related processes influencing the timing and direction of positions on NAFTA, three models are estimated. The first model treats the two equations for timing and direction separate. The second model uses the SUDCD Weibull estimator derived above where the parameter for correlation of the errors of the discrete choice and duration models is constant. The third model allows this correlation to be different for Republican and Democratic representatives. The relationship between timing and position direction is potentially stronger for Democrats as the President was a pro-NAFTA Democrat and therefore likely to apply pressure on members of Congress from his own party to vote in favor of NAFTA and/or offer inducements and side-payments to do so. This provides Democratic representatives with a greater incentive to hold out on declaring a position to see if their votes are critical to the final outcome and/or to receive greater

[5] One new independent variable, *net endorsements* is added by Boehmke (2006).

Table 9.2 Independent Variables in Models of NAFTA Position Timing and Direction

Independent Variable	Description	Expected Effect On Direction	Expected Effect On Timing
Union Membership	Level of unionization in a representative's district, measured as the percentage of workforce that is unionized	–	+ (if high or low)
Mexican Border	Records whether or not a representative's district shares a border with Mexico	+	+
Perot Vote	Support in the representative's district for Ross Perot (known for his strong anti-NAFTA stance) in the 1992 U.S. presidential election	–	+ (if high or low)
Household Income	Median household income in the representative's district, divided by 10,000	+	+ (if high or low)
Corporate Contributions	Share of representative's total campaign contributions accounted for by business interests	+	+
Labor Contributions	Share of representative's total campaign contributions accounted for by labor interests	–	+
NAFTA Committee	Representative is a member of a congressional committee that took up the matter of membership in NAFTA		+
Democratic Leadership	Representative is part of the Democratic congressional leadership		+/–
Republican Leadership	Representative is part of the Republican congressional leadership		+
Party Affiliation	Dummy variable coded 1 if the representative is a Democrat; 0 if Republican	+	
Ideology	Representative's ideological position on a conservative-liberal scale based on his/her congressional voting record	+	+
Net Endorsements	Difference between number of representatives that have already declared in favor of NAFTA and those that have declared against	+	

For timing, the expected effect is on the hazard rate; thus a "+" indicates an earlier declaration of a position.

side-payments from the President. Furthermore, greater pressure and side-payments from the President are likely to cause Democrats to vote in favor of NAFTA. Results for all three models are presented in Table 9.3.

Table 9.3 Separate and SUDCD Weibull Models of Timing and Direction of Positions on NAFTA

	Separate		Weibull SUDCD[6]		Weibull SUDCD	
Covariate	Estimate	s.e.	Estimate	s.e.	Estimate	s.e.
Vote						
Labor Contributions (net %)	-2.929***	(0.621)	-2.913***	(0.615)	-2.938***	(0.615)
Mexican Border	0.521	(0.600)	0.571	(0.598)	0.606	(0.597)
Union Membership	-6.089***	(1.477)	-6.143***	(1.465)	-6.046***	(1.466)
Household Income	2.749**	(1.084)	2.721**	(1.080)	2.783**	(1.081)
Democrat	-0.558***	(0.211)	-0.568***	(0.209)	-0.584***	(0.210)
Net Endorsements	0.011**	(0.005)	0.010*	(0.005)	0.010**	(0.005)
Constant	0.825***	(0.177)	0.793***	(0.177)	0.806***	(0.177)
Timing						
Corporate Contributions	0.140**	(0.068)	0.144**	(0.068)	0.150**	(0.068)
Labor Contributions	-0.116	(0.077)	-0.112	(0.077)	-0.112	(0.077)
Mexican Border	-0.220***	(0.039)	-0.220***	(0.039)	-0.217***	(0.039)
Democratic Leadership	-0.023	(0.030)	-0.023	(0.030)	-0.022	(0.030)
Republican Leaderhsip	-0.071**	(0.032)	-0.072**	(0.031)	-0.072**	(0.032)
NAFTA Committee	-0.004	(0.014)	-0.004	(0.014)	-0.003	(0.014)
Ideology	0.005	(0.017)	0.005	(0.017)	0.005	(0.017)
Union Membership	-0.352**	(0.146)	-0.337**	(0.145)	-0.328**	(0.145)
Union Mem. * Ideology	0.486**	(0.236)	0.449*	(0.234)	0.459*	(0.235)
Household Income	0.035	(0.108)	0.041	(0.107)	0.043	(0.107)
Income * Ideology	-0.021	(0.016)	-0.022	(0.015)	-0.021	(0.015)
Constant	6.059***	(0.018)	6.058***	(0.018)	6.057***	(0.018)
Correlation ($Z^{-1}(\alpha)$)						
Intercept			0.365	(0.230)	0.075	(0.338)
Democrat					0.480	(0.463)
Duration dependence ($ln(p)$)	2.086***	(0.0436)	2.089***	(0.0435)	2.090***	(0.0435)
N	434		433		433	

Estimates for duration equations have a time-to-failure interpretation.
$***p < 0.01, **p < 0.05, *p < 0.1$

We see that there is very little difference in the estimates for the observed covariates, and their statistical significance, across the models. The only notable difference is that the interaction term between union membership and ideology is statistically significant when the duration model is estimated separately from the discrete-choice model at the 0.05 level but is only significant at the 0.1 level in the combined SUDCD models.

Of most interest here is the estimation of the correlation between the stochastic elements of the discrete choice model of NAFTA positioning and the duration model of position timing. The constant correlation is parameterized in the log-likelihood function by $Z^{-1}(\alpha)$. Fisher's Z transformation is employed, as the correlation must be between -1 and 1, such that:

$$\alpha = Z(\alpha^*) = (\exp(2\alpha^*) - 1)/(\exp(2\alpha^*) + 1) \tag{9.21}$$

and the correlation parameter $\rho = \alpha/4$. The SUDCD estimator reports $Z^{-1}(\alpha) = \alpha^*$. Thus to calculate $\hat{\rho}$, when not differentiating between representatives according to party affiliation, one would substitute the reported value for the intercept (0.365) into equation (9.21) for α^* to generate a value for α, which when divided by four gives a correlation coefficient of 0.087.

The parameterization of the correlation changes when we wish to calculate different different values of $\hat{\rho}$ for Democratic representatives and Republican representatives. When this is done, the parameterization of the correlation is:

$$\alpha^* = Z^{-1}(\alpha) = \alpha_0^* + \alpha_1^* \times \text{Democrat}_i \tag{9.22}$$

with α_0^* equal to the reported intercept (0.075) and α_1^* equal to the value reported for Democrat (0.480). Substituting these values into equation (9.22), we find that $\hat{\rho}$ for Republican members of Congress equals 0.019 and for Democrats it equals 0.126. The overall estimated parameter is 0.56 (0.48 + 0.075) with a χ^2 value of 2.88 and a p-value of 0.09.

Substantively, the results indicate several things. First, that the correlation between the errors in the two models is positive and statistically significant at the 0.1 level, indicating that the unobservable influences on NAFTA vote choice and the timing of that choice are positively related. Those unobservable factors such as side-payments and competing political pressures that caused legislators to hold out longer before declaring their position, also caused them to vote in favor of NAFTA. Furthermore, these factors had a much greater impact on the Democratic members of Congress than on Republicans as is evident by the comparatively small value of $\hat{\rho}$ for Republican representatives. This accords with Boehmke's expectations. Democrats, more than Republicans, faced competing pressures from their party, and the President, to approve NAFTA while their constituents lobbied them to reject the agreement. As the final vote on NAFTA was close, Democrats that held out until as late as possible to see if their vote would be crucial were forced to cast their vote in favor of NAFTA to make sure the measure passed. If the vote did not appear to be close, they could have voted against NAFTA and thus neither angered their support base nor the President. Furthermore, the positive relationship between a late declaration of position and taking a pro-NAFTA stance supports the argument that an unobservable process whereby indifferent Democratic members held out on taking a position in order to extract side-payments from NAFTA supporters, and the President, before agreeing to cast their votes in favor of NAFTA.

9.6 Promising Future Directions

This review has focused on the modeling of systems of equations using one or more duration equations as a technique for uncovering the effects of latent variables on our outcomes of interest, which in the event history context is typically the time until an event occurs. In the extant literature, these models are generally based on SUR or SEQ approaches which jointly estimate equations that model interdependent processes (e.g. time to event). However, Hays and Kachi (2009) have gone one step further to consider interdependence of durations between actors (or units), which means that the time to a particular event for one actor depends on the time to the same event for other actors. They provide several persuasive examples which suggest that accounting for unobservable inter-unit interdependence is important. For instance "the time it takes for states to enter wars [often] depends on the time it takes other states to make these decisions" (2), and the time for states to decide on a policy issue, such as allowing casino gambling, may depend on when other states have adopted similar policies. Similarly, the decision to lower gas prices may depend on when competitors do the same. While Hays and Kachi (2009) argue that their estimation approach applies to both interdependence between units and interdependence between times to events, the former has previously been modeled with a spatial duration model with correlated errors (Darmofal 2009) and with a spatial lag model (Honoré & de Paula, in press), both of which may be of interest to readers.

Note that the literature on estimating duration models in systems of equations is dominated by parametric approaches for the duration equations. In contrast, most of the single equation duration literature in the social and behavioral sciences is dominated by the use of the Cox model. The advantage of relaxing the distributional assumption about the time until an event occurs suggests that incorporation of the Cox model into a simultaneous setup would be promising.

Combinations of different types of duration models and simultaneous equations is also promising, particularly, the use of the competing risks duration model, which allow for more than one type of event. Problems such as the duration of education and the time to form a union where a union can be cohabitation or marriage would be an example of two durations where one of them (union formation) requires the use of a competing risks duration model. Similarly, the competing risks of the duration of cohabitation where cohabitation can end with either marriage or break-up and the discrete choice model for fertility may prove to be a useful model.

In short, the area of simultaneous equations and duration models is a flourishing area of research with wide applicability to key questions in the social sciences. Recent modeling developments have provided new questions as well as new answers to old questions.

Appendix: Computer Code

Example: Duration and Discrete Choice in the NAFTA Study

The results for the NAFTA study examining interdependence between duration and discrete choice models were generated using STATA. Below is code, adapted from Boehmke's own do-files, that allows for replication of his analysis for the models presented here.

Installing the estimation program SUDCD:

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

Defining the likelihood functions for the discrete exponential estimator:

```
program define expdisc
version 7
args lnf theta1
quietly replace ‘lnf’ = ln(exp(-exp(-‘theta1’)))
  if $ML_y1==1
quietly replace ‘lnf’ = ln(1-exp(-exp(-‘theta1’)))
  if $ML_y1==0
end
```

Separate estimation of the discrete choice model of NAFTA support using the discrete exponential estimator:

```
ml model lf expdisc (vote = contdiff mexbordr pscenter
  hhcenter partyid numdiff)
ml search
ml maximize

predict ystar_exp if e(sample), xb
generat yhat_exp = exp(-exp(-ystar_exp))
recode yhat_exp 0/0.5=0 0.5/1=1
tab yhat_exp vote, matcell(crosstab)
```

Separate estimation of the duration model of NAFTA position timing:

```
stset timing, failure(position)

streg corptpct labtpct mexbordr dleader rleader
  ncomact ideol pscenter
inter1 hhcenter inter2, d(weibull) time
```

Estimating the combined SUDCD models:
(Right-censoring is hard coded into the likelihood functions, which requires explicitly declaring the _rtcens variable using a dummy variable, rtcsensr, which is set to one if the observation is right censored.[7])

```
gen _rtcens = rtcensr
```

Joint estimation of the duration and discrete choice equations without estimating different coefficients of correlation for subsets of the sample:

```
sudcd timing corptpct labtpct mexbordr dleader rleader
  ncomact ideol pscenter inter1
hhcenter inter2, discrete(vote= contdiff mexbordr pscenter
  hhcenter partyid numdiff)
dist(weibull) time rtcensor(rtcensr)
```

Joint estimation of the duration and discrete choice equations, estimating different coefficients of correlation for Democrats and Republicans. The variable name for party affiliation is *partyid*:

```
sudcd timing corptpct labtpct mexbordr dleader rleader
  ncomact ideol pscenter inter1
hhcenter inter2, discrete(vote= contdiff mexbordr
  pscenter hhcenter partyid numdiff)
dist(weibull) time rtcensor(rtcensr) rho(partyid)

display "Correlation for Republicans (rho): "
((exp(2*([Z_alpha]_b[_cons]))-1)/
  (exp(2*([Z_alpha]_b[_cons]))+1))/4

display "Correlation for Democrats (rho): "
((exp(2*([Z_alpha]_b[_cons] + [Z_alpha]_b[partyid]))-1)
  /(exp(2*([Z_alpha]_b[_cons] + [Z_alpha]_b[partyid]))+1))/4

test [Z_alpha]partyid + [Z_alpha]_cons = 0
```

Acknowledgements We thank Aya Kachi and Kentaro Fukumoto for correspondence and discussion.

[7] To run without right-censoring, just set this variable equal to zero.

References

Aalen, O. O. (1975). *Statistical inference for a family of counting processes*. PhD thesis, University of California, Berkeley.

Adair, L. S., Popkin, B. M., & Guilkey D. K. (1993). The duration of breast-feeding: How is it affected by biological, sociodemographic, health sector, and food industry factors? *Demography, 30,* 63-80.

Addison, J. T., & Portugal, P. (1989). Job displacement, relative wage changes, and duration of unemployment. *Journal of Labor Economics, 7,* 281–302.

Allison, P. D. (1984). *Event history analysis: Regression for longitudinal event data*. Thousands Oaks: Sage.

Baizán, P., Aassave, A., & Billari, F. C. (2004). The interrelations between cohabitation, marriage and first birth in Germany and Sweden. *Population and Environment, 25,* 531-561.

Billari, F. C., & Philipov D. (2004). Women's education and entry into a first union: A simultaneous-hazard comparative analysis of Central and Eastern Europe. *Vienna Yearbook of Population Research,* 91-110.

Boehmke, F. J. (2006). The influence of unobservable factors in position timing and content in the NAFTA vote. *Political Analysis, 14,* 421-428.

Boehmke, F. J., Morey, D. S., & Shannon, M. (2006). Selection bias and continuous-time duration models: Consequences and a proposed solution. *Journal of Political Science, 50,* 192-207.

Box-Steffensmeier, J. M., & Jones, B. S. (2004). *Event history modeling: A guide for social scientists.* New York: Cambridge University Press.

Box-Steffensmeier, J. M., Arnold, L. W., & Zorn, C. J. W. (1997). The strategic timing of position taking in Congress: A Study of the North American Free Trade Agreement. *American Political Science Review, 91,* 324-338.

Boyle, P. J., Kulu, H., Cooke, T., Gayle, V., & Mulder, C. H. (2008). Moving and union dissolution. *Demography, 45,* 209-222.

Collett, D. (2003). *Modelling survival data in medical research.* London: Chapman & Hall.

Coppola, L. (2004). Education and union formation as simultaneous processes in Italy and Spain. *European Journal of Population, 20,* 219-250.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, B, 34,* 187-220.

Cox, D. R. (1975). Partial likelihood. *Biometrika, 62,* 269-276.

Darmofal, D. (2009). Bayesian spatial survival models for political event processes. *American Journal of Political Science, 53,* 241-257.

Dwivedi, T. D., & Srivastava, V. K. (1978). Optimality of least squares in the seemingly unrelated regression equation model. *Journal of Econometrics, 7,* 391-395

Fleming, T. R., & Lin, D. Y. (2000). Survival analysis in clinical trials: Past developments and future directions. *Biometrics, 56,* 971-983.

Fukumoto, K. (2009). *What happens depends on when it happens: Continuous or ordered event history analysis.* Working paper: Faculty of Law, Gakushuin University, Tokyo.

Golub, J., & Collett, D. (2002). Institutional reform and decision making in the European Union. In M. Hosli & A. van Deemen (Eds.), *Institutional Challenges in the European Union*. London: Routledge.

Grambsch, P. M., & Therneau, T. M. (1994). Proportional hazards tests and diagnostics on weighted residuals. *Biometrika, 81,* 515-526.

Hays, J. C., & Kachi, A. (2009). *Interdependent Duration Models in Political Science.* Paper presented at the Annual Meeting of the American Political Science Association, Toronto, Sept. 3-6, 2009.

Honoré, B., & de Paula, A. (in press). Interdependent durations. *Review of Economic Studies*.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation for incomplete observations. *Journal of the American Statistical Association, 53,* 457-481.

Larsen, U., & Vaupel, J. W. (1993). Hutterite fecundability by age and parity: Strategies for frailty modeling of event histories. *Demography, 30,* 81-102.

Lillard, L. A. (1993). Simultaneous equations for hazards. *Journal of Econometrics, 56,* 189-217.

Maitra, P., & Sarmistha P. (2007). *Birth spacing, fertility selection and child survival: Analysis using a correlated hazard model.* Institute for the Study of Labor (IZA), Discussion Paper No. 2878.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Report, 50,* 163-170.

Oakes, D. (2001). Biometrika centenary: Survival analysis. *Biometrika, 88,* 1.

Olshanksy, S. J., & Carnes, B. A. (1997). Ever since Gompertz. *Demography, 34,* 1-15.

Pindyck, R. S., & Rubinfeld, D. L. (1991). *Econometric models and economic forecasts.* New York: McGraw-Hill.

Quiroz Flores, A. (2008). *Copula functions and bivariate distributions for survival analysis: An application to government survival.* NYU Department of Politics. Working paper.

Rosholm, M., & Svarer, M. (2001). Structurally dependent competing risks. *Economics Letters, 73,* 169-173.

Singer, J. D., & Willett, J. B. (1993). It's about time: Using time survival analysis to study duration and the timing of events. *Journal of Educational Statistics, 18,* 155-195.

Therneau, T. M., Grambsch, P. M., & Fleming, T. R. (1990). Martingale based residuals for survival models. *Biometrika, 77,* 147-160.

Therneau, T. M., & Grambsch, P. M. (2001). *Modeling survival data: Extending the Cox model.* New York: Springer-Verlag.

Tuma, N. B. (1976). Rewards, resources, and the rate of mobility: A nonstationary multivariate stochastic model. *American Sociological Review, 41,* 338-360.

van Montfort, K., Oud, J., & Satorra, A. (Eds.) (2004). *Recent developments on structural equation models: Theory and Applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association, 57,* 348-368.