11

Strategies for Temporal Dependence

Duration Models

Up to this point in the book we have derived likelihood functions and evaluated models under the assumption that the Y_i are conditionally independent, i.e., $\Pr(y_1, y_2, \ldots, y_n|\boldsymbol{\theta}) = P(y_1|\theta_1) \times P(y_2|\theta_2) \times \ldots P(y_n|\theta_n)$. This are many threats to this assumption. For example, observations close together in space may all be influenced by some neighborhood-specific factor, such as a leaky nuclear reactor. This is known as spatial dependence. Dependence can also arise when events or observations take place over time.

There are several ways of thinking about temporal dependence. For example, we can consider frequency rather than time (Beck, 1991). Another approach considers specific values or events in a temporal sequence. In such time series we have repeated observations of some unit i at fixed intervals $t, t+1, \ldots$. We are concerned that there might be correlation across them, often referred to as serial (auto)correlation (Hamilton, 1994). Time series models, such as an autoregressive integrated moving average (ARIMA) and many others, typically organize the data to examine their characteristics in the time domain. There is a vast and highly developed literature on these topics. We do not pursue them here because that would require an entire volume. Instead, we concentrate on a third approach to temporal dependence: modeling the time between or until discrete events.

11.1 INTRODUCTION

Social scientists are frequently concerned with how long things last and how often they change. How long will one nation be at war with another? How long is a leader's tenure in office? How long after an election until a government is formed? How long can we expect a worker to remain unemployed? Analysis of this sort of data goes by different names across disciplines. In political science and sociology, models of these processes are most commonly called

https://doi.org/10.1017/9781316888544.014 Published online by Cambridge University Press

event history models; in economics they are referred to as *duration* models; in demography and health sciences the models are referred to as *survival* models; and in engineering they are called *failure time* or *hazard* models. Whatever they are called, these approaches are concerned with modeling what we will call *spells*, i.e., the length of time an observation spends in a particular condition (typically called a state) before either it transitions to different condition or the study period ends.

The basic idea behind duration models is to assume that spell durations can be viewed as a random variable, T. This permits the use of a probability density for a spell of length t. In turn, we can model the expected value of that distribution as a function of covariates, which may also vary over time.

The definition of what constitutes being in a spell is akin to a categorical variable. The most basic duration models employ a binary categorical variable, but more complex models for multiple, competing risks and the like have analogues in ordered or multinomial categorical data. The major wrinkle that justifies the development of an entirely new class of models for duration data is the problem of *censoring*.

As an example, consider a population of countries. We are concerned with the length of time a country spends at war. When observing conflict, we end up with two types of observations: those for which the duration of the conflict is known, i.e., we observe the entire spell from beginning to end, and those for which the duration of conflict is unknown because the observation period ends before the conflict is observed to end. Event history models account for both types of observations in the likelihood function.

In this chapter, we concentrate on duration models, binary time-series cross-section (BTSCS) data, and the semi-parametric Cox proportional hazards model. We then develop an overview of parametric event history models, such as the Weibull, on our way to a specialized form, the split-population event history model. The Cox model has rapidly become the standard in most empirical applications due to its less-restrictive assumptions about the baseline hazard rate and its flexibility in incorporating more complicated functional forms.

11.2 DATA STRUCTURES AND THINKING ABOUT TIME

More than most analysis problems, categorical time series data require us to think ahead about the types of models we are likely to fit before we set up our data in a format to be analyzed. Two key issues are (1) whether we consider time discrete or continuous and (2) whether to use the spell as the unit of analysis as opposed to some putatively natural unit of time such as a year. The latter decision will largely be driven by whether we wish to include time-varying covariates, i.e., explanatory variables that change through time – perhaps within a spell – in our models. Most social science data are collected in fixed intervals, such as years, not in terms of episodes that have a beginning and an end.

11.2.1 Discrete and Continuous Time

If we imagine transitions from one state to another can occur at any arbitrary moment, then we are thinking in continuous time. If events can only occur within specific intervals, then time is discrete. Discrete time is made from continuous time. The world we actually inhabit is one of continuous time, and few political processes are constrained to occur at specific discrete intervals. Nevertheless, actual measurement regularly coarsens time into discrete units or observational periods (days, weeks, years, etc.). Put another way, time is continuous, but our data are almost always discrete. Many texts make much of the distinction between discrete time and continuous time models. Our take is that this is less of a concern since our data are almost always discrete and grouped. The practical question involves the length of spells relative to our temporal precision in observing transitions from one state to another. The longer the spell relative to the time aggregation unit, the closer we are to observing continuous time. For example, if the average spell length is a year and a half and our data are at the annual level, then we clearly have a discrete process. If instead we have daily data, we observe transitions that are relatively finely grained moments in time.

If our observation of transitions is fine-grained, then it makes sense to consider time to be continuous, modeling the spell as the unit of analysis. Why? Recall the models of conflict that we covered in Section 4.5.1: for most dyads, in most years there is no conflict. We saw that most models for discrete processes were not terribly good at predicting both conflict and nonconflict. The challenge here is that there are so many nonconflict episodes relative to conflictual ones, i.e., our data are "zero-inflated." A highly accurate model is one that simply produces 0, since there are so few 1s. Similarly, if we observe a survival process at the daily level and most spells are years long, then we will have many time periods in which no events occur. It makes sense, then, to consider the length of the spell as our outcome of interest, especially if we have covariates that are not changing within a spell. Alternatively, if we have data aggregated to a high level relative to the length of a spell, then a discrete time approach is probably called for.

11.2.2 Organizing Data as a Spell-Level

Event history analysis (in continuous time) takes all data to be organized into spells. Table 11.1 gives an example of how this might look, taken from King et al.'s (1990) paper on cabinet durations. The key novelty here is the censoring variable, indicating whether a unit survived until the end of the observation period. Note that all covariates are constant within spells.

TABLE 11.1 *Spell organization of data.*

Case	Duration in Months	X_1	X_2	X_3	X_4	Censored?
1	7.00	93.00	1.00	1.00	5.00	Yes
2	27.00	62.00	1.00	1.00	1.00	Yes
3	6.00	97.00	1.00	1.00	1.00	Yes
4	49.00	106.00	1.00	1.00	1.00	No
5	7.00	93.00	1.00	1.00	4.00	Yes
⋮	⋮	⋮	⋮	⋮	⋮	
314	48.00	60.00	0.00	1.00	1.00	No

11.2.3 Continuous Time with Time-Varying Covariates: Counting Process

If we want to include covariates that may change within a spell, then we need to set up data in a different format, referred to as a *counting process*, due to Andersen and Gill (1982). Setting up data as a counting process is useful for more-complicated event history models, including those for time-varying covariates, repeated, and multiple events. The first step is to define the *risk set*, i.e., the units and time periods at risk for the transition or event of interest. We must then identify the intervals at which the covariates change value and the times when the events of interest occur. The counting process is formulated as a [start,stop] interval. The censoring variable indicates that a unit has exited the risk set, either due to an event of interest or because the observational period ended.

Table 11.2 displays some of the data used in Ahlquist (2010b), organized as a counting process. The `status` variable indicates whether the event of interest occurs in that period.

The counting-process construction in Table 11.2 reflects stacked time series data, also known to some as time-series cross-section (TSCS), and to others as a panel.[1] The primary difference from here is in modeling strategies. The BTSCS (discrete time) approach, discussed in the next section, uses `status` as the dependent variable and the number of periods since an event occurred for that unit as a measure of time. Models developed under the assumption of continuous time require the counting process (`start` and `stop`) in addition to the censoring indicator.

Before turning to this, we present briefly the data structure of duration models in Figure 11.1. The top row of this illustration is a vector of numbers

[1] The distinction between panel and TSCS data depends on whether n (the number of units) is much greater than the number of time periods, T. When $n >> T$ we are in a panel world. $T >> n$ is referred to as TSCS data. The distinction is not hard-and-fast and derives from whether a particular estimator relies on asymptotics in T or n to derive its properties. See Beck and Katz (1995).

TABLE 11.2 *Counting process data, from Ahlquist (2010b).*

Index	id	Country	Event Time	Start	Stop	Status	tsle.ciep	lag.infl	lag.unemp	growth	enpp
1	AUS1974Q1	Australia	1	0	1	0	36.00	16.38	2.13	3.79	2.52
2	AUS1974Q2	Australia	2	1	2	0	44.00	16.55	2.10	3.79	2.52
3	AUS1974Q3	Australia	3	2	3	0	4.00	15.90	2.11	3.79	2.52
4	AUS1974Q4	Australia	4	3	4	0	12.00	15.25	2.64	3.79	2.52
⋮		⋮		⋮		⋮		⋮		⋮	
36	AUS1982Q4	Australia	36	35	36	0	65.00	9.12	7.02	2.92	2.64
37	AUS1983Q1	Australia	37	36	37	1	74.00	8.54	8.72	−3.00	2.23
⋮		⋮		⋮		⋮		⋮		⋮	
2,381	USA1999Q1	United States	101	100	101	0	12.50	1.29	4.43	3.39	1.99
2,382	USA1999Q2	United States	102	101	102	0	25.00	1.44	4.29	3.39	1.99
2,383	USA1999Q3	United States	103	102	103	0	37.50	1.62	4.25	3.39	1.99
2,384	USA1999Q4	United States	104	103	104	0	50.00	1.80	4.25	3.39	1.99

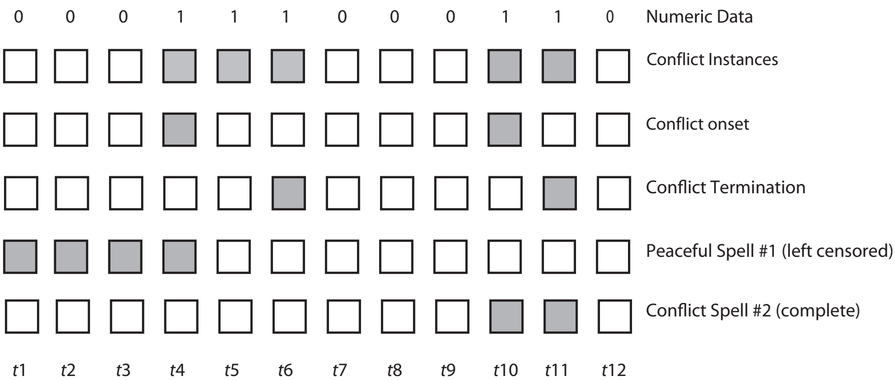


FIGURE 11.1 An illustration of duration data using conflicts, onsets, durations, spells.

that represent whether a particular country in the sample set being analyzed has a conflict during a particular year. Much conflict data are organized as a collection of such vectors in BTSCS format. The second row is a visual representation of the same information, with filled squares representing conflicts, and empty ones shown for years in which there was no conflict in that country. The third row uses shading to identify the years in which there was an observed transition from peace to conflict whereas the fourth row identifies conflict termination–years of transition from conflict to peace. The row labeled `peaceful spell #1` shows that the first four time periods make up the first spell of non-conflict until (and including) the year in which the peace ends and the conflict begins. This spell is left-censored because we do not observe when it begins. The final row encodes the years consisting of the second spell of conflict, one in which we observe both the onset and termination. Standard analyses often focus on the numeric data, but duration models focus on explaining the lengths of the spells; Standard analyses have covariates that are measured for each t, but duration models are really concerned, if at all, with what the covariates look like during the spells.

11.3 DISCRETE TIME: THE BTSCS APPROACH

The discrete time BTSCS setup closely mirrors the analysis of panel and time series, cross-section data with continuous response variables. BTSCS data represent repeated observations of a unit i over time, t. The outcome of interest is a simple dichotomous variable, Y_{it}, which takes on a value of 1 if the event of interest occurred for unit i in period t and 0 otherwise. Typically we have n units, each of which is observed for T periods. The data for each unit are stacked on top of one another, producing an outcome vector of length nT. Beck et al. (1998) point out that the aggregation of events by time induces a discrete (i.e., grouped) structure in time. For example, if we have yearly observations on law

https://doi.org/10.1017/9781316888544.014 Published online by Cambridge University Press

adoption by US states, we only observe whether a state adopts the law in a year; all adoptions within a year are considered equivalent.

Beck et al. (1998) argue that fitting a logit, probit, or complementary log-log model to this pooled, stacked data is both appropriate and equivalent to event history modeling so long as the analyst includes either (1) indicator variables for event time or (2) (cubic) splines in event time.[2] The event time indicators that Beck et al. have in mind are $T-1$ dummy variables for the number of periods since that unit last experienced an event. These time dummies have the direct interpretation of a "baseline hazard," i.e., the probability of an event occurring in any particular interval given that it has not yet occurred and before we consider covariates. It turns out this is equivalent mathematically to a duration model, developed in the next section.

Formally, the logit model with time indicators is

$$
\begin{aligned}
Y_{it} &\sim f_B(y_{it}; \theta_{it},) \\
\theta_{it} &= \text{logit}^{-1}\left(\mathbf{x}_{it}^{\intercal}\boldsymbol{\beta} + \tau_t \mathbb{1}_{t_i}\right),
\end{aligned}
$$

where $\mathbb{1}_{t_i}$ is a dummy variable that equals one whenever unit i has gone t periods without an event and zero otherwise.

This BTSCS approach is simple to implement when the data are organized as a panel, rather than as a collection of episodes. BTSCS models can be interpreted along the lines of simple binary GLMs. But these models do require explicit description of temporal dependence across observations in the form of time indicators or splines. Carter and Signorino (2010) highlight two problems here. First, splines are complicated and often implemented incorrectly, in addition to being difficult to interpret. Second, a common problem with the indicator-variable approach is the small number of observations with very high times-to-failure. These small numbers can induce perfect separation. Carter and Signorino (2010) propose using a simple cubic term in event time $(\tau_1 t_i + \tau_2 t_i^2 + \tau_3 t_i^3)$ rather than splines or time dummies to get around these exact problems. In our experience a simple linear trend often works well.

The BTSCS approach does have weaknesses. It requires time-varying data but, at the same time, it does not require explicit consideration of censoring. The BTSCS approach does not direct attention to the time between events or how covariates may affect that interval, focusing instead only on the event occurrence. While the BTSCS approach continues to see use, there are also flexible and easy-to-estimate duration models to which we turn.

[2] Splines are piecewise polynomials, with the pieces defined by the number of "knots," or locations in which the values of the two polynomials on either side are constrained to be equal. Temporal splines have the advantage of using many fewer degrees of freedom (one per knot) than indicator variables for each period, but the number and location of the knots is not an obvious decision. We do not take up splines further here. We also note that there are many other ways of addressing this degrees of freedom problem than just splines.

11.4 DURATION MODELS

The combination of spells and censoring motivate the duration model. But it often turns out that it is mathematically convenient to model spells in terms of the *hazard rate* rather than durations. In this section we survey the basic pieces of duration models: the density function over t, the *survivor function*, the *hazard function*, and the *cumulative* (or integrated) *hazard function*. All the pieces of event history analysis – the density, survivor function and hazard rate – are mathematically dependent. Defining one determines the rest.

The terms "survival," "hazard," "failure," and "risk" are artifacts of these models' development in the context of time-to-failure analysis in engineering and patient survival in the health sciences. While these terms lend the discussion something of a sinister air, we emphasize that they are purely technical terms and imply no normative judgment about whether being in a spell is good or bad. To be clear, a "failure" or "event" simply refers to the end of a spell or, equivalently, a transition from one state to another, e.g., from war to peace or life to death. Similarly, a unit is "at risk" for failure if it is still in a spell, i.e., it is at risk for a transition from one state to another.

11.4.1 Survivor and Hazard Functions

Suppose we have n observational units indexed by i. For each unit we observe either the complete length of its spell or the length of its spell until the observation period is truncated. More formally, an observed failure time $T_i = \min\{Y_i^*, C_i^*\}$, where Y_i^* is some latent failure time and C_i^* is a latent censoring time.

For brevity we suppress subscripts until discussing the likelihood. Let T be a random variable representing spell duration. It therefore takes values in $[0, \infty)$. We can represent the probability that $T = t$, i.e., that i's spell is of length t, by specifying a density function $f(t)$, with corresponding distribution function $F(t)$.

Definition 11.1 (Survivor and hazard functions). Let $F(t) = \Pr(Y_i^* \leq t)$. We define the *survivor function*, $S(t)$:

$$S(t) = 1 - F(t) = \Pr(T > t).$$

The survivor function describes the probability that i's spell will be at least of length t. The survivor function allows us to specify the *hazard function*, $h(t)$:[3]

$$h(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)}. \tag{11.1}$$

[3] Formally, the hazard function is defined as a limit: $h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$, making it easier to see that the hazard function is a conditional distribution.

https://doi.org/10.1017/9781316888544.014 Published online by Cambridge University Press

The hazard function describes the probability that a spell will end by t given that it has survived until t. The value of the hazard function at t is referred to as the *hazard rate*. Empirically, the hazard rate is simply the proportion of units still at risk for failure at t that fail between t and $t+1$. We can sum the hazard rates in continuous time to express the "accumulation" of hazard. This function, $H(t)$, is called the cumulative or integrated hazard rate.

Theorem 11.1 (Integrated hazard rate).

$$H(t) \equiv \int_0^t h(x)dx = -\log S(t).$$

Proof We know that

$$\begin{aligned} f(t) &= \frac{d}{dt}F(t) = \frac{d}{dt}(1 - S(t)) \\ &= -\frac{d}{dt}S(t). \end{aligned} \tag{11.2}$$

Substituting Equation 11.2 into 11.1 we obtain

$$h(t) = \frac{-\frac{d}{dt}S(t)}{S(t)}.$$

So,

$$\begin{aligned} \int_0^t h(x)dx &= \int_0^t \frac{-\frac{d}{dx}S(x)}{S(x)}dx \\ &= -(\log S(t) - \log S(0)) \\ &= -\log S(t), \end{aligned}$$

where the second line results from an application of the chain rule: $\frac{d}{dx}\log f(x) = \frac{f'(x)}{f(x)}$. □

11.4.2 A Likelihood

The problem in constructing a likelihood for survival data is that we generally do not observe failure times for all units. Units for which our observation window ends before we observe failure are *right-censored*. Units that begin their risk exposure prior to the beginning of the observation period are called *left-censored*. The models discussed herein are capable of incorporating both left- and right-censoring but, for brevity, we focus on the far more common right-censored case. These observations contribute information on survival up to the censoring point; they do not contribute any information about failure. Thus, if we define a censoring indicator δ_i, which equals 1 if i is censored and 0 otherwise, the generic form of the likelihood is

https://doi.org/10.1017/9781316888544.014 Published online by Cambridge University Press

$$\mathcal{L} = \prod_{i=1}^{n} f(t_i)^{1-\delta_i} S(t_i)^{\delta_i}. \tag{11.3}$$

Different parametric duration models derive from different distributional assumptions for $f(t)$. The exponential, Weibull, log-logistic, and log-normal models are the most common. Each has different implications for the hazard rate and its relationship to time, i.e., "duration dependence." The exponential model is a special case of the Weibull and assumes a baseline hazard rate to be flat with respect to time. In this model the risk of an event is *time independent*. The Weibull distribution implies a model that fits a baseline hazard that is monotonic (increasing, decreasing, or unchanging) though time. Log-normal and log-logistic models allow for hazard rates to be unimodal, i.e., increasing and then decreasing in time. All these models can be estimated and evaluated with all the standard tools and procedures associated with MLE.

11.4.3 Writing Down a Duration Model

There are two main ways of expressing duration models: as "accelerated failure time" (AFT) and the proportional hazards specification. The former is more common in fields that use parametric event history models, whereas the latter is more common in the social sciences. The difference between the two is that AFT models link covariates to survival times, while hazard models link covariates to the hazard rate. But since specifying a hazard function also specifies a survival function, the two are very similar. Parameter estimates from an AFT model are simply -1 times the parameter estimates from a hazard rate specification.

Accelerated Failure Time (AFT)

In AFT, models covariates act multiplicatively on the failure time. Thus, AFT models focus on the rate at which an observation continues in the current state.

The model is given:

$$S(t|X) = \psi\left(\frac{\log(t) - \mathbf{x}_i^{\intercal}\boldsymbol{\beta}}{\sigma}\right),$$

where ψ can be any standard survival distribution and σ is a scaling parameter. This can be rewritten:

$$\begin{aligned} \log(T_i) &= \mathbf{x}_i^{\intercal}\boldsymbol{\beta} + \sigma u_i \\ &\Updownarrow \\ T_i &= \exp(\mathbf{x}_i^{\intercal}\boldsymbol{\beta} + \sigma u_i), \end{aligned}$$

which directly models the survival times, T_i. The only question is what distributional assumption to impose on u_i.[4] For example, if we assume that

[4] σ can be factored out if it is constant for each observation.

https://doi.org/10.1017/9781316888544.014 Published online by Cambridge University Press

$u_i = \frac{1}{p}\epsilon_i$ and that ϵ_i is Type-I extreme value distributed, then we have the Weibull model. If we impose the constraint that $p = 1$, the exponential model results.

Proportional Hazards

The more common way of writing down duration models is in the hazard rate form. The exponential and Weibull models, as well as the Cox model described below, are often expressed in terms of hazard rates:

$$h_i(t) = h_0(t)\exp(\mathbf{x}_i^\top \boldsymbol{\beta}), \tag{11.4}$$

where $h_0(t)$ is the baseline hazard, possibly a function of time (but not covariates), and common to all units. The specification for h_0 determines the model. For example, if $h_0(t) = pt^{p-1}$, then we have the Weibull model; with $p = 1$ we again have the exponential.

In case you were wondering ... 11.1 The Weibull distribution

Waloddi Weibull was a Swedish mathematician who described a two-parameter distribution now known as the Weibull distribution, even though Maurice Fréchet developed it much earlier. Fréchet published some of his papers in Esperanto.

We say $T \geq 0$ follows the Weibull distribution with parameters $\lambda > 0, p > 0$:

$$T \sim f_W(t; \lambda, p) = \begin{cases} \frac{p}{\lambda}\left(\frac{t}{\lambda}\right)^{p-1}\exp\left(-(\frac{t}{\lambda})^p\right) & t \geq 0, \\ 0 & t < 0. \end{cases}$$

We refer to λ as the scale parameter and p the shape parameter. $\mathrm{E}[T] = \lambda + \Gamma(1+p^{-1})]$ and $\mathrm{var}(T) = \lambda^2[\Gamma(1+2p^{-1}) - \Gamma(1+p^{-1})^2]$. Note that when $p = 1$ the Weibull distribution reduces to the exponential distribution.

Given the Weibull density we can express the hazard and survivor functions as

$$h_W(t) = p\lambda^p t^{p-1},$$
$$S_W(t) = \exp(-(\lambda t)^p).$$

The distribution gives the time to failure in the sense that the failure rate is an exponential function of the length of time. Values of the shape parameter, p, bifurcate at 1; below that threshold the failure rate is decreasing, whereas above 1, it grows. Note that pt^{p-1} is the baseline hazard for a Weibull proportional hazards model.

The $\exp(\mathbf{x}_i^{\top}\boldsymbol{\beta})$ expression represents unit-level differences that scale the baseline hazard up or down. In models that have the proportional hazards property, the intercept term is frequently omitted since it applies equally to all units and is therefore part of the baseline hazard rate. The multiplicative change in the hazard ratio – $\exp(\beta)$ – implies that β represents the change in the log hazard ratio. Positive coefficients imply increasing hazards (shorter survival times), whereas negative coefficients imply decreasing hazards (increasing survival times). In proportional hazards models it is often easier to interpret $\exp(\beta)$, since these are understandable in terms of proportional changes in the hazard ratio. For example, if $\exp(\hat{\beta}_1) = 1.4$, then a unit increase in x_1 increases the hazard rate by a factor of 1.4, or 40%.

From Equation 11.4 we can adduce an interpretation for the regression parameters β as well as demonstrate the *proportional hazards* property. To see both, suppose that we have only one covariate, x_1, and we want to compare the hazard rates for two different units, i and j, with covariate values of x_{i1} and x_{j1}, respectively.

$$\frac{h_i(t, x_1 = x_{i1})}{h_j(t, x_1 = x_{j1})} = \exp[(x_{i1} - x_{j1})\beta_1].$$

In comparing the hazard rates, the baseline hazard cancels out, which implies that the *hazard ratio* is fixed across time, a property referred to as proportional hazards. Put another way, proportional hazards implies a model where the "effect" of a covariate on the hazard rate is the same no matter when in a spell the covariate might change its value. The proportional hazards assumption is conceptually similar to the BTSCS assumption that the $\boldsymbol{\beta}$ are fixed through time and constant across observations. Note that the proportional hazards property is an *assumption* in several types of event history models, but it is one that can be tested.

11.5 THE COX MODEL

The preceding discussion begs the question of which among the many parametric models to use. Should the baseline hazard rate be decreasing, increasing, or fluctuating? Rarely is theory any guide. To the extent we fit models to test propositions, our ideas are generally about the relationship between covariates and the outcome. So the choice of baseline hazard is data-driven. We are then left with two starkly opposing solutions. The first, which we mention in passing, is to fit a variety of models and evaluate them on the basis of parsimony, predictive performance, or some other heuristic. The second is to refuse to specify the baseline hazard altogether, leaving the form of time dependence unspecified as a consequence. Pursuing this second option is the domain of the Cox proportional hazards model.

This Cox proportional hazard model can be expressed as

$$h_i(t) = h_0(t)\exp(\mathbf{x}_i^\top \boldsymbol{\beta}),$$

but while the parametric models assume a specific form for h_0, the Cox model leaves it unspecified. Since part of the model is left unspecified the Cox model is often referred to as "semi-parametric."

11.5.1 The Partial Likelihood

Without a baseline hazard, we cannot fully specify a likelihood function. But if we assume that the times between events contribute nothing to our understanding of the relationship between the covariates and the hazard rate, then we can use the ordered failure times to construct a *partial likelihood*, which has all the standard properties of a likelihood, and therefore can be maximized.

To derive the partial likelihood, assume for the moment that we have n observational units for which we observe n spells. Of these spells, u of them are uncensored, i.e., we observe the event. We can order the failure times as $t_1 < \ldots < t_u$. Note that these are all strict inequalities; we assume (for the moment) that there are no events occurring at the same moment. We then model the probability that an event happens to unit i at time t_j, conditional on there being an event at time t_j. The probability of an event happening to i, given that i has not yet failed, is simply the hazard rate for i. The probability of there being an event at t_j is simply the sum of all the hazards for all the units at risk for failure at t_j. If we denote R_j to be the risk set at time t_j, then the conditional probability can be expressed as

$$\begin{aligned}\frac{h_i(t_j)}{\sum_{l\in R_j} h_l(t_j)} &= \frac{h_0(t_j)\exp(\mathbf{x}_i^\top \boldsymbol{\beta})}{\sum_{l\in R_j} h_0(t_j)\exp(\mathbf{x}_l^\top \boldsymbol{\beta})}\\ &= \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta})}{\sum_{l\in R_j}\exp(\mathbf{x}_l^\top \boldsymbol{\beta})}.\end{aligned}$$

Thus the partial likelihood for the Cox model is given by

$$\mathcal{L}_P = \prod_{i=1}^{n}\left[\frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta})}{\sum_{l\in R_j}\exp(\mathbf{x}_l^\top \boldsymbol{\beta})}\right]^{1-\delta_i},$$

where, again, $\delta_j = 1$ if the case is right-censored. Comparing this expression to equation 11.3, there is an important difference: in the Cox partial likelihood the censored cases contribute information about the cases at risk for an event, but they contribute nothing to the estimation of failure times. This partial likelihood can be maximized with numerical techniques.

Tied Events

The derivation of the Cox model assumed that no two units failed in the same time interval. Indeed, if we were really observing events in continuous time, then the chances of two events actually occurring at the same instant are vanishingly small. However, data are aggregated in time slices (seconds, weeks, years). In practice ties occur regularly.

One of the great strengths of the Cox model is its flexibility in handling tied events. Several different methods have emerged to adjust the partial likelihood to accommodate ties. They go by the names *Breslow*, *Efron*, *averaged* or *exact partial*, and *exact discrete* or *exact marginal*. The Breslow method is the default method for most statistical packages *but not in $\mathcal{R}$*. It is generally held to be the least accurate, with this problem becoming more severe the more ties there are. The Efron method tends to work better (and is the default in $\mathcal{R}$). The two "exact" methods are the most precise, but can impose severe computational burdens if there are a lot of ties.

Residuals from Cox (and Other) Models

The concept of the empirical model residual becomes more complicated outside of the simple regression context. In event history models, it is even more complex because we have censored observations and, in the case of the Cox model, a semi-parametric model. It turns out that there are three different types of "residuals" that each have different uses. We outline their uses here, omitting technical discussions of their origins and properties:

Cox-Snell
: Cox-Snell residuals are used to examine overall model fit. They should be distributed as unit exponential. Residual plots allow us to visually examine whether this is (approximately) true for a particular model.

Schoenfeld
: Schoenfeld residuals are used for evaluating the proportional hazards property. Box-Steffensmeier and Jones (2004) note that Schoenfeld residuals can "essentially be thought of as the observed minus the expected values of the *covariates* at each failure time." (2004:121, emphasis added)

Martingale
: Martingale residuals are the most intuitive to think about but perhaps the most mathematically complex. They are given by the observed censoring indicator minus the expected number of events, as given by the integrated hazard rate. Martingale residuals are $\widehat{M}_i = \delta_i - \hat{H}_0(t_i)\exp\left(\mathbf{x}_i^{\mathsf{T}}\hat{\boldsymbol{\beta}}\right)$, where δ_i is an indicator of the event for observation i, and $\widehat{H}_0(t_i)$ is the estimated cumulative hazard at the final time for observation i. Martingale residuals are plotted against included (or possibly excluded) covariates to evaluate whether they are included appropriately.

https://doi.org/10.1017/9781316888544.014 Published online by Cambridge University Press

Deviance Deviance residuals are Martingale residuals rescaled so as to be symmetric about 0 if the model is "correct."

Score Score residuals are useful for identifying high-leverage observations.

11.5.2 An Example: Cabinet Durations

We illustrate some elements of event history analysis in $\mathcal{R}$ using the cabinet duration data from King et al. (1990). To begin, we load the `survival` library. In `survival` there are three basic steps to fitting a model:

1. Set up the data appropriately (spell-level or counting process).
2. Create a survival object using `Surv()`.
3. Fit the model as normal using the survival object as the response variable.

The King et al. (1990) data are organized at the spell level, containing 314 government cabinets in parliamentary democracies. The outcome variable is the duration of the government in months; covariates include a slate of variables capturing various aspects of the party system and political climate. Typical results are displayed in Table 11.3. Note the reporting of both $\hat{\beta}$ and $\exp(\hat{\beta})$.

What about the proportional hazards assumption? Figure 11.2 plots the Schoenfeld residuals against time; if the proportional hazards assumption holds, the slope of this relationship should be 0. There is no compelling visual evidence that the assumption is violated here.

These plots of residuals are helpful but largely subjective. A more formal test of the relationship between the residuals and time is available and presented

TABLE 11.3 *Cox model of government cabinet duration.*

	$\hat{\beta}$	$\exp(\hat{\beta})$	$\sigma_{\hat{\beta}}$
Majority Government	−0.49	0.62	0.13
Investiture	0.56	1.74	0.14
Volatility	−0.00	1.00	0.00
Polarization	0.03	1.03	0.01
Fractionalization	0.00	1.00	0.00
Crisis	−0.01	0.99	0.00
Formation Attempts	0.11	1.12	0.05
Opposition Party	−0.06	0.94	0.42
n	314		
$\log \mathcal{L}$	−1,293		
AIC	2,601		
BIC	2,630		

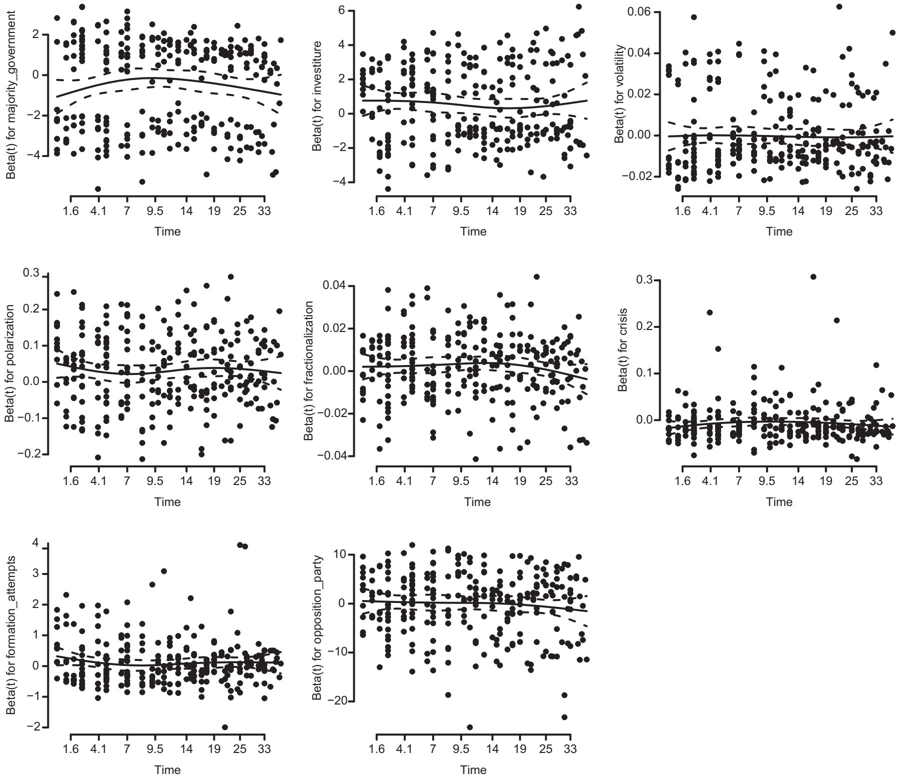


FIGURE 11.2 Schoenfeld residuals against adjusted time.

in Table 11.4. In none of the variables are we able to reject the null of no relationship. There is no reason to believe the proportional hazards assumption is violated here.

Interpreting Results

A quick glance at the (exponentiated) coefficients in the BUTON implies that, for example, requiring a vote of investiture increases the hazard of coalition termination by about 74%. Coalitions controlling a legislative majority have about a 38% longer survival time than those in the minority, all else constant. But, as usual, the nonlinear form of the model makes it difficult to envision what we mean by "all else constant." If we want to interpret the model on something like the scale of the dependent variable, more work is needed. We can construct meaningful scenarios and then generate model predictions on an interpretable scale.

TABLE 11.4 *Formal tests of the correlation of residuals with time.*

	ρ	χ^2	p-Value
Majority Government	0.00	0.01	0.94
Investiture	−0.05	0.52	0.47
Volatility	−0.01	0.05	0.81
Polarization	−0.01	0.01	0.89
Fractionalization	−0.06	0.86	0.35
Crisis	0.01	0.07	0.79
Formation Attempts	−0.03	0.26	0.61
Opposition Party	−0.06	1.23	0.26
GLOBAL	NA	2.93	0.93

Hazard rates are often difficult to understand or convey to audiences for two reasons: they are conditional, and the sign of estimated coefficients in a proportional hazards model is the opposite of its implications for survival. Presenting results on the survival scale is often easier to comprehend. Figure 11.3 displays the expected survival times for governments depending on whether the cabinet requires a vote of investiture. We see that only about 35% of governments requiring investiture are expected to survive for at least twenty months; the rate is about 55% for those not needing such a vote. Put another way, we expect about half of governments requiring an investiture vote to have fallen at around 15 months, compared to about 22 months for those not requiring a vote. This particular plot leaves off estimates of uncertainty around these predictions, but our standard simulation techniques are capable of generating them.

Code Example 11.1 produces the Cox analysis and graphics.

11.5.3 Pros and Cons of the Cox Model

The Cox model has much to recommend it. It avoids explicit modeling of duration dependence. It is flexible in its ability to handle ties, repeated events, and other complications. It has a relatively straightforward proportional hazards interpretation. Unless you care about the duration dependence directly, then Cox is often the model of choice.

The Cox model also has disadvantages. It avoids any need to model duration dependence and relies only on the ordering information, ignoring the duration between events. One month is the same as 50 months. The Cox model also has a tendency to overfit the data, since the baseline hazard is completely determined by the proportion of observations surviving in a period. That is, it is highly tuned to the data at hand. As a result, those focusing on out-of-sample prediction and extrapolation (e.g., engineers) often use parametric

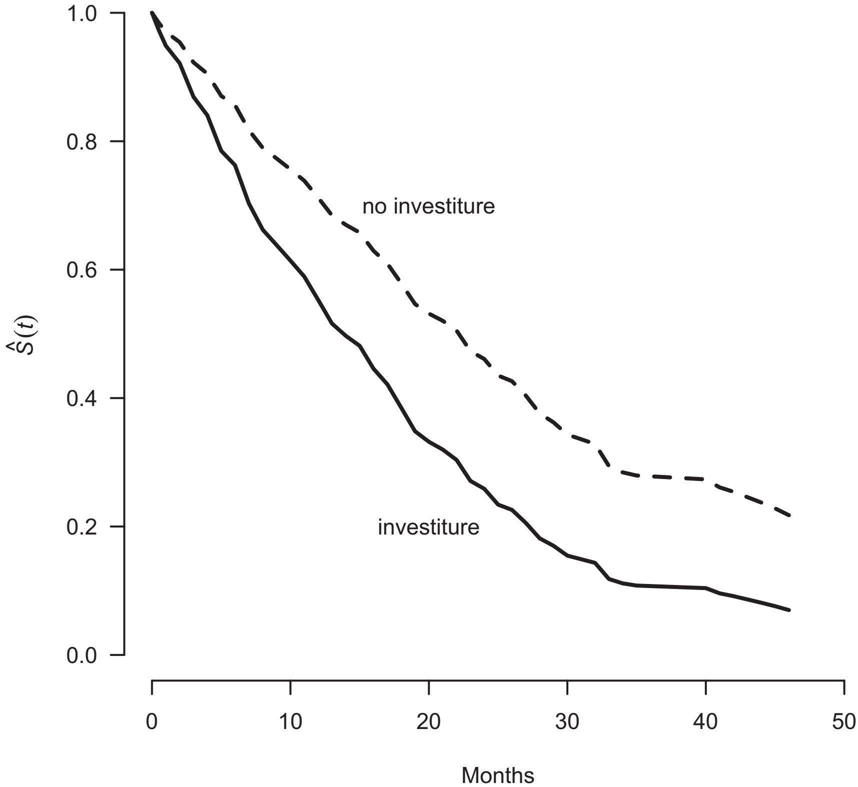


FIGURE 11.3 Expected government survival times depending on whether the cabinet requires a vote of investiture, with all over covariates held at central tendencies.

models. Extracting interpretation on the level of the dependent variable, often a good thing to do, can be quite difficult in more-complicated Cox models with stratified variables, competing risks, or "frailty" terms.[5] Finally, like most duration models it treats all observations as being determined by the same forces. New Zealand and the Ukraine would be subject to the same influences in terms of a model of civil conflict. It would seem that the baseline level of risk in the former is much less than in the latter. We now turn to a way to model this.

[5] Stratification in a Cox model estimates a separate baseline hazard rates for different values of the stratified variable. Competing risks models are those that model multiple types of transitions at once. "Frailty" models are a form of random coefficient or "random effects" Cox model in which observational units are heterogeneous in the propensity to survive.

https://doi.org/10.1017/9781316888544.014 Published online by Cambridge University Press

$\mathcal{R}$ Code Example 11.1 ***Fitting and interpreting Cox PH models***

```
library(survival)
library(foreign)
sdat <- read.dta("coalum2.dta")
cox.fit <- coxph(Surv(time=duration, event=censor12) ~ majority_government +
            investiture + volatility + polarization + fractionalization +
            crisis + formation_attempts + opposition_party, data=sdat)
cox.zph(cox.fit) #testing the proportional hazards assumption
deets <- coxph.detail(cox.fit)
times <- c(0,deets$time)  #observed failure times
h0 <- c(0,deets$hazard) #hazard evaluated a sample mean of covariates
x0 <- c(0)-cox.fit$mean
h0 <- h0*exp(t(coef(cox.fit))%*%x0) #baseline hazard
x.inv <- c(1,1,median(sdat$volatility), #investiture scenario
  mean(sdat$polarization),
  mean(sdat$fractionalization), mean(sdat$crisis),
  median(sdat$formation_attempts),mean(sdat$opposition_party))
x.ninv <- x.inv
x.ninv[2] <- 0  #no investiture scenario
h.inv <- h0*exp(t(coef(cox.fit))%*%x.inv)
h.ninv <- h0*exp(t(coef(cox.fit))%*%x.ninv)
Sinv <- exp(-cumsum(h.inv))   # survival function
Sninv <- exp(-cumsum(h.ninv)) # survival function
plot(times,Sinv, type="l", xlab="Months", lwd=2,
  ylab=expression(hat(S)(t)), bty="n", las=1,
  ylim=c(0,1), xlim=c(0,50))
lines(times,Sninv, lwd=2, lty=2)
text(x=20, y=c(0.2, 0.7), labels=c("investiture", "no investiture"))
```

11.6 A SPLIT-POPULATION MODEL

Suppose you are interested in modeling how long leaders stay in power, and whether the transition from one leader to the next is by regular, constitutional means, or by some irregular process, such as a coup d'état, revolution, or succession crisis. Consider the irregular transfer of power to be the event of interest. To model the timing until the onset of a problem event, we use a split-population duration approach. Initially developed in a health sciences context to examine the survival of medical patients under the assumption that some are "cured" and others remain at risk, split-population models assume a heterogeneous group of subjects. We can use them here to model the survival of polities, given some risk of irregular leader transition, war, or a similar type of failure event. This approach involves mixing different distributions, something we discussed in the context of the zero-inflated Poisson model (see Section 10.5.3).

The hypothesis that not every polity is at risk of irregular regime change motivates the split-population approach. For all practical purposes, countries like Germany, Canada, or Japan are highly unlikely to experience irregular transitions within the time period we consider here. However, many countries in

https://doi.org/10.1017/9781316888544.014 Published online by Cambridge University Press

Africa and the Middle East experienced irregular transitions in the past decade. The important point from a modeling perspective is identifying countries, or in this case country-months, which are "at risk" of failure. From a policy perspective these are the places where estimating an expected duration until the next event makes sense.

Figure 11.4 illustrates the intuition behind this approach. As shown in the left panel, there are two types of polities. First are those that may have had an event but essentially are immune from further events (Country B). These include countries that never had any events but are not shown in this illustration. The second type of polity is at risk for future events (Country A). The split-population approach models first the separation of locations into type A or B, denoted by the *if* in the panel. The next part of the model determines the duration of time until the next event, denoted by *when*. The right panel illustrates the differences in base hazard rates under the assumption that all locations have the same risk profile (the standard Weibull) compared to the baseline risk that assumes the population of locations consists of two types: those at risk and those immune from risk.

The basic likelihood of this kind of situation may be thought of as a mixture of two distributions: a Bernoulli distribution determining whether a unit is at risk and then a second distribution describing duration. We define the variable $\pi_i(t)$ to be an indicator variable equal to one if t_i is part of a spell that ultimately ends in an observed event of interest and 0 otherwise. We can build the model as

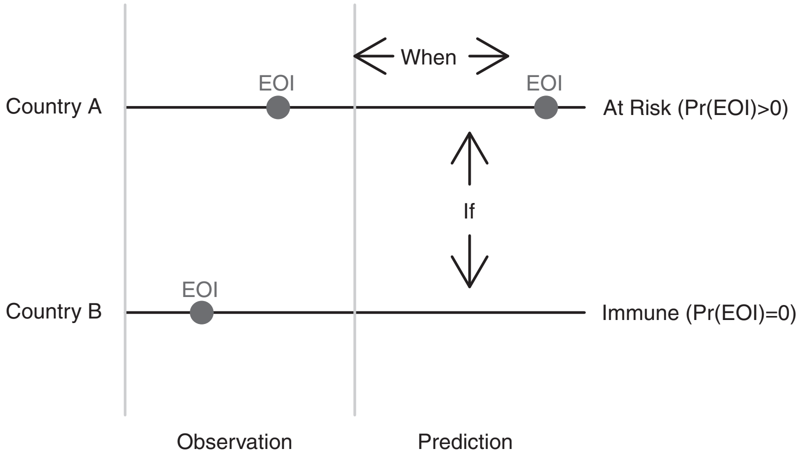
$$\begin{aligned} T_i &= \min\{Y_i^*, C_i^*\}, \\ \pi_i &\sim f_B(\theta_i), \\ \theta &= \text{logit}^{-1}(\mathbf{z}_i^\mathsf{T}\boldsymbol{\gamma}), \\ T_i &\sim f(t_i; \lambda), \\ \lambda &= \exp(-\mathbf{x}_i^\mathsf{T}\boldsymbol{\beta}), \\ S(t_i|\mathbf{x}, \mathbf{z}) &= \theta_i S(t) + (1 - \theta_i). \end{aligned}$$

Building on Equation 11.3, we can derive the likelihood as a product of the risk and duration:

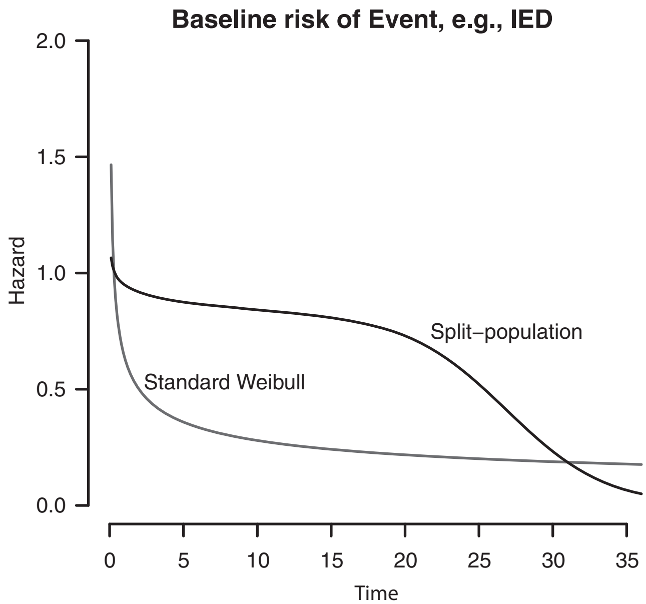
$$\mathcal{L} = \prod_{i=1}^{n} \left[\theta_i f(t_i)\right]^{1-\delta_i} \times \left[(1-\theta_i)S(t_i)\right]^{\delta_i},$$

where $f(t_i)$ is the failure rate at time t_i, $S(t_i)$ is the survival function, and δ_i is the indicator of right-censoring. The split-population model is set up for two populations, one of them at risk for an event, the other "immune."

This likelihood function reflects a mixture of two equations: a first step classifying risk and immunity, and a second step describing expected duration in a spell. One advantage of this modeling approach is that it allows covariates to have both a long-term and a short-term impact, depending on whether they' appear in the **z** or the **x** vector. Variables that enter the at-risk equation



(a) Splitting Population



(b) Baseline Hazard

FIGURE 11.4 Country A is at risk; Country B is not at risk. Mixing these two yields a risk assessment that is too low while overestimating the risk decline. EOI refers to the Event Of Interest.

have a long-term impact because they change the probability of being at risk at all. Variables in the second duration equation can be thought of as having a short-term impact that modifies the expected duration until the next failure.

https://doi.org/10.1017/9781316888544.014 Published online by Cambridge University Press

To complete the model we must choose a distribution function for $f(t)$. The $\mathcal{R}$ package `spduration` (Beger et al., 2017) currently admits two: the Weibull and log-logistic. The Weibull density allows for hazard rates that are increasing, constant, or decreasing over survival time, while the log-logistic density can fit rates that have a peak at a particular survival time.

Note that the model focuses on whether the observational unit at time t is in the risk set – e.g., a country at a particular point in time – not the unit *per se*. As the covariates change over time, so can our estimate of whether an observation is in the risk pool at that point in time. So, while it is helpful on a conceptual level to speak of observational units as being at risk or immune, in a technical sense we should refer to them at a specific point in time. In an analysis of countries, Canada may not be at risk in 2014, but it may be at risk in 2015 if some unanticipated disaster leads to conditions that we associated with country-months in the risk pool.

In case you were wondering ... 11.2 The Log-logistic distribution

Let $X \sim f_L(x;\mu,\sigma)$. If $T = \exp(X)$, we say T follows the *log-logistic* distribution, sometimes called the *Fisk* distribution. The log-logistic distribution has several different parameterizations. The one implemented in `spduration` uses scale parameter $\lambda > 0$ and shape paramter $p > 0$:

$$T \sim f_{lL}(t;\lambda,p) = \begin{cases} \frac{p\lambda(\lambda t)^{p-1}}{(1+(\lambda t)^p)^2} & t \geq 0, \\ 0 & t < 0. \end{cases}$$

Given the log-logistic density we can express the hazard and survivor functions as

$$h_{lL}(t) = \frac{p\lambda(\lambda t)^{p-1}}{1+(\lambda t)^p},$$

$$S_{lL}(t) = \frac{1}{1+(\lambda t)^p}.$$

To use the observation-time data with a split-duration model, we need, among other variables, a counter for each spell that indicates the number of months since the start time or previous failure time. This introduces an issue when we do not observe the previous failure time, i.e. left-censoring. For example, in a study of irregular regime changes, if we were to start all counters in 2001, the United Sates in 2004 would have the same counter value as Serbia, which had an irregular exit in October 2000 (Slobodan Milosevic). To mitigate this problem, we use a much earlier date as the start date for building the duration-related variables, if at all possible.

https://doi.org/10.1017/9781316888544.014 Published online by Cambridge University Press

Survival data consist of two types of spells, those which end in an event of interest and those that are right-censored. The key assumption with the split-population approach involves the coding of spells or country-months as susceptible ($\pi_{it} = 1$). The split-duration approach "retroactively" codes $\pi_{it} = 1$ if period t is part of a spell that ended in an observed event in unit i. Spells that are right-censored take on $\pi_{it} = 0$, as do spells that end when an observation leaves the data set with no event taking place in the last spell (e.g., an observation ceases to exist). Treating right-censored spells as "cured" can be problematic, since they may later, after we observe more data, end in failure. The probabilistic model for π_i partially mitigates this by both incorporating the length of a censored spell and sharing information across cases known to be at risk as well as those coded as cured.

11.6.1 Interpretation of Split-Duration Estimates

Once estimated, there are several different quantities we can calculate from a split-population duration model. We focus on the *conditional hazard rate*, which describes the probability that a country will experience failure during a given time period, given that it has already survived without failure to that point, and considering that some countries will never experience failure. In fewer words, it is the split-duration model's best guess of a particular unit's failure during the time period in which we are interested. To fully understand this, we can break the conditional hazard rate for a country at a time t down into two components: the unconditional hazard rate and the probability that it is at risk of failure at all.

The *unconditional hazard rate* is the probability that an observation will experience a failure in a given time frame (say a month), given that it has survived without failure up until that month. We can estimate it based on historical data on previous events in other observations and previous events in the same observation. Thus the unconditional hazard rate for a particular month gives the probability of an event, given that it will have been some number of months at that point since the previous event occurred in that observation.

There are three things to note about the hazard rate. First, it is specific to a given observation and a given time since the last event. The hazard rate for events in one observation is likely to be different from the hazard rate our model estimates for another. Similarly, the hazard rate for six months from the last event will probably be different from the hazard rate at twelve months from the last event. Second, these changes in the hazard rate over time can follow specific shapes. They can be flat, meaning that the hazard of an event does not change over time. This is the case, for example, for the decay of radioactive elements. But the hazard rate for events could also have a bump in the beginning, indicating that these are more likely a short time after the last event, but less likely over time. In any case, a variety of shapes are possible. The third thing to note is that this unconditional hazard rate assumes that all observations are subject to experiencing the events. That hardly seems sustainable, since we know that some observations are immune to the particular risk under study. To that end, a split-population duration model also tries to group observations that are at risk separately from those that are effectively "cured."

The *risk probability* is an estimate that a given observation at a specific time falls into either the susceptible group ($\pi = 1$) or the cured group ($\pi = 0$). It is an estimate *at a given time* because it also depends on how much time has passed since the last event. To calculate the conditional hazard, we combine our estimated hazard rate with the estimated probability that a country at a given time is susceptible to an event. In other words, *conditional hazard* = unconditional hazard $\times$ risk probability. More formally, given the density distribution and estimated parameters we are interested in, the conditional hazard $h(t, \theta)$, where both the at-risk probabilities and hazard are conditional on survival to time t:

$$\theta(t) = \frac{1-\theta}{S(t) + (1-\theta)(1-S(t))}, \tag{11.5}$$

$$h(t,\theta) = \frac{f(t,\theta)}{S(t,\theta)} = \frac{\theta(t)f(t)}{(1-\theta(t)) + \theta(t)S(t)}. \tag{11.6}$$

Equation 11.6 shows that the conditional risk rate is decreasing over event time because, as time passes, the surviving cases increasingly consist of the immune $(1-\theta)$ that will never fail. In Equation 11.6 for the conditional hazard, the failure rate in the numerator is conditional on the probability that a case is in the risk set, given survival up to time t. The denominator is an adjusted survivor function that accounts for the fraction of cured cases by time t: $(1-\theta(t))$.

We estimate the probability of an event in (say) February 2014 to be the unconditional probability of that event when it has been so many (say, 23) months since the last event times the probability that the observation is in the "at risk" group given its characteristics and given that it has been 23 months since the last event. In this way we can get a probability estimate for an event that takes into account the changing hazard of events over time but which also corrects for the fact that some observations will never experience an event.

The coefficient estimates from a split-population model can be interpreted similar to how we viewed ZIP coefficients. The coefficients in the risk equation are logistic regression parameters that indicate whether a change in a variable increases or decreases the probability that a country-month is in the set of susceptible unites; exponentiated coefficients indicate the factor change in risk probability associated with a one-unit change in the associated variable. The duration part of the model is in AFT format, and for interpreting them it is convenient to think of the dependent variable as being survival time, or, equivalently, time to failure. A negative coefficient shortens survival and thus hastens failure (higher probability of an event at time t), while a positive coefficient prolongs survival and thus delays failure (lower probability of an event at time t).

11.6.2 An Example

In order to illustrate practical implementation of a split-population model, we adapt the Beger et al. (2017) reexamination of Belkin and Schofer (2003). Belkin and Schofer model coups d'état. They explicitly distinguish long-term structural risk factors for coups from short-term triggering causes that can explain the timing of a coup in an at-risk regime. The data we analyze include 213 coups. As examples, we fit a conventional Weibull model and Weibull and log-logistic split-population duration models, including Belkin and Schofer's index of structural coup risk in the risk equation. Table 11.5 is the requisite BUTON reporting coefficients from the duration equation and then estimates from the risk equation.[6] The duration models are in accelerated failure time format, and the coefficient estimates are on the log of expected time to failure. The negative coefficient for military regimes, for example, means that the expected time to a coup is shorter in military regimes than nonmilitary regimes, holding all other factors constant. In the risk equation, positive coefficients mean a higher risk of coup. Thus, military regimes have a higher risk of experiencing a coup. Looking at the AIC and BIC, we see that the split-population model outperforms the Weibull AFT model. The split-population models are indistinguishable, so we will continue to focus on the log-logistic form.

A plot of the conditional hazard is the probability of a coup at a time t, conditional on the covariates in the risk and duration equations and survival up to time t. We fix covariates at their sample means. These are shown in Figures 11.5 and 11.6. Figure 11.6 compares the conditional hazard with covariates held at mean values (panel A) and when covariates are set to a high-risk, military-regime values (panel B). The conditional hazard is much higher and steeper in B than in A, reflecting the increased risk of coup.

Code for estimating these models appears as Code Example 11.2.

Out-of-Sample Testing

One of the strengths of a parametric model is its ability to generate out-of-sample forecasts. In this case we use data from 1996 onwards as the test set, and prior data for training purposes. For the training data, we need to subset the training set first, so that coups in the test set do not influence the risk coding in the training data. For the test set, we add the duration variables and *then* subset the test set. Since the test set is later in time than the training set, there is no contamination of the risk coding, but if we subset the data before building the duration variables, we will start all duration counters at 1996, when in fact we can safely use the previous historic coup information.

[6] The AFT parameterization of the Weibull model does not report an intercept.

https://doi.org/10.1017/9781316888544.014 Published online by Cambridge University Press

TABLE 11.5 *Weibull and split-population Weibull and log-logistic regression models of coups, 1960–2000.*

		Split Population	
Duration Model	**Weibull**	**Weibull**	**log-Logistic**
Intercept		3.22	2.40
		(0.17)	(0.21)
Instablity	−0.06	−0.09	−0.09
	(0.01)	(0.01)	(0.02)
Military regime	−2.33	−1.55	−1.13
	(0.17)	(0.19)	(0.21)
Regional conflict	6.03	5.08	−2.52
	(2.44)	(2.72)	(2.16)
log p	−0.15	0.03	−0.45
	(0.05)	(0.05)	(0.06)
Risk Model			
Intercept		−0.44	2.93
		(3.89)	(1.87)
Risk index		1.65	0.59
		(0.83)	(0.32)
log GDPpc		0.35	−0.36
		(0.68)	(0.28)
Military regime		11.57	10.82
		(3.92)	(9.29)
Recent war		−2.19	−0.53
		(1.97)	(0.94)
Regional conflict		−5.30	−5.43
		(12.49)	(5.62)
South America		−0.55	2.10
		(2.18)	(1.45)
Central America		−1.02	−0.39
		(1.41)	(0.73)
n	4,250	4,250	4,250
Num. events	213	213	213
log $\mathcal{L}$	−704	−662	−662
AIC	1,417	1,330	1,331
BIC	1,442	1,349	1,350

The default prediction in `spdur` is the conditional hazard, which can be thought of as a predicted probability. As a result, we can use the same tools for visualizing the performance of survival models that we used for other forms of binary classifiers, including the separation plot. Figure 11.7 displays the out-of-sample performance of both the Weibull and log-logistic models.

https://doi.org/10.1017/9781316888544.014 Published online by Cambridge University Press

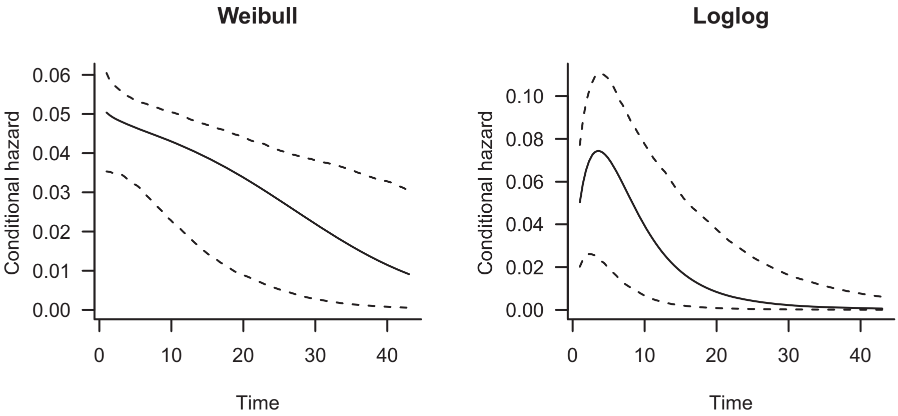


FIGURE 11.5 Conditional hazard rates for the split-population Weibull and log-logistic model of coups with all covariates held at sample means.

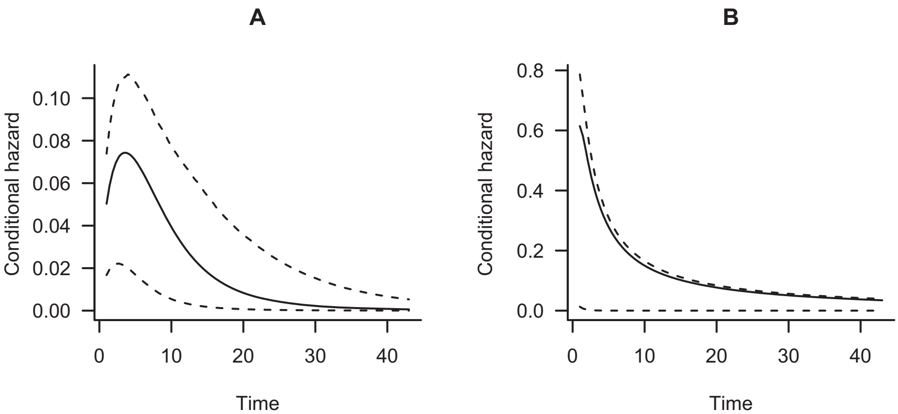


FIGURE 11.6 Plots of the hazard rate for the log-logistic model of coups. The left graph uses the default mean values for covariates, while graph B uses user-specified variable values for a high-risk military regime.

Code for conducting the out-of-sample forecasting exercise appears in Code Example 11.3.

11.7 CONCLUSION

This chapter introduced a framework for using likelihood principles when there is temporal dependence between observations. But rather than focusing on the serial approach that emphasizes repeated measurements over time, we focused on duration models that explicitly model time itself. We have developed the likelihoods and illustrated how they may be modified to address a variety of

https://doi.org/10.1017/9781316888544.014 Published online by Cambridge University Press

$\mathcal{R}$ Code Example 11.2 *Split-population duration models*

```
data(bscoup)  #in spdur library.
bscoup$coup <- ifelse(bscoup$coup=="yes", 1, 0)
bscoup <- add_duration(bscoup, "coup", unitID="countryid",
   tID="year", freq="year", ongoing = FALSE) #formatting for duration model
weib_model <- spdur(
  duration ~ milreg + instab + regconf,
  atrisk ~ couprisk + wealth + milreg + rwar + regconf +
  samerica + camerica, data = bscoup, silent = TRUE)

loglog_model <- spdur(
  duration ~ milreg + instab + regconf,
  atrisk ~ couprisk + wealth + milreg + rwar + regconf +
  samerica + camerica, data = bscoup,
  distr = "loglog", silent = TRUE)

weib_aft <- aftreg(Surv(time=t.0, time2=duration, event=coup) ~
  milreg + instab + regconf, dist="weibull", param = "lifeExp",
  data=bscoup[rownames(bscoup)%in%rownames(weib_model$mf.dur),])
plot_hazard(weib_model, main = "A",bty="n",las=1)
plot_hazard(loglog_model, main = "A",bty="n",las=1)
plot_hazard(loglog_model,
            xvals = c(1, 1, 10, 0.05),
            zvals = c(1, 7, 8.64, 1, 1, 0.05, 0, 0),
            main = "B",bty="n",las=1)
```

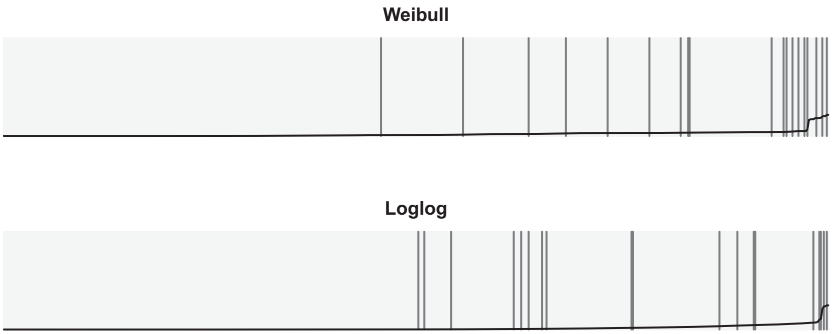


FIGURE 11.7 Out-of-sample separation plots.

questions – including population heterogeneity in the form of split-population models. In evaluating these models, we showed how hazard rates can be viewed as predicted probabilities, bringing us full circle from BTSCS to spells and back to binary data.

This chapter is far from a complete or canonical treatment of duration models. By focusing on intuition around temporal dependence and connect to likelihood principles and tools – including split populations and forecasting – we ignored other complications. One of these is left- and interval-censoring. Left-censoring means that an event happened prior to the beginning of data collection, but it is not known exactly when. Interval-censoring is most often

https://doi.org/10.1017/9781316888544.014 Published online by Cambridge University Press

$\mathcal{R}$ **Code Example 11.3** ***Evaluating duration models out-of-sample***

```
coup_train <- bscoup[bscoup$year < 1996, ]  #training set
coup_train <- add_duration(coup_train, "coup", unitID = "countryid",
    tID = "year", freq = "year", ongoing = FALSE) #formatting for spdur
coup_test <- add_duration(bscoup, "coup", unitID = "countryid", tID = "year",
                           freq = "year", ongoing = FALSE)
coup_test  <- coup_test[coup_test$year >= 1996, ]
weib_model2   <- spdur(
    duration ~ milreg + instab + regconf,
    atrisk ~ couprisk + wealth + milreg + rwar + regconf + samerica +
      camerica,
    data = coup_train, silent = TRUE)

loglog_model2 <- spdur(
    duration ~ milreg + instab + regconf,
    atrisk ~ couprisk + wealth + milreg + rwar + regconf + samerica +
      camerica,
    data = coup_train, distr = "loglog", silent = TRUE)

weib2_test_p   <- predict(weib_model2, newdata = coup_test)
loglog2_test_p <- predict(loglog_model2, newdata = coup_test)
obs_y <- coup_test[complete.cases(coup_test), "coup"]
library("separationplot")
par(mfrow=c(2,1),mar=c(2,2,2,2))
separationplot(weib2_test_p,    obs_y, newplot = FALSE)
separationplot(loglog2_test_p, obs_y, newplot = FALSE)%\vspace*{-24pt}
```

ignored because it occurs so frequently in social sciences: an event occurred but exactly when in the interval is not recorded. Most country-year data are interval-censored. The basic issue is that the models assume that censoring is not endogenous, i.e., that it is not caused by the impending event.

We have also surveyed only a fraction of many distributional assumptions that can be made in duration analysis. Many others are used, including the log-normal, extreme value distributions as well as piecewise functions. Random effects or "frailty" are another way to approach heterogeneity, allowing individual parameters to vary by observation, in a way similar to random effects in other models. These are useful when there are recurrent events or when there are strata or clusters of observations that may be affected by similar factors.

11.8 FURTHER READING

Applications

Jones and Branton (2005) compare the BTSCS approach to the Cox model in the study of policy adoption across US states. Thrower (2017) uses the Cox model to investigate the duration of executive orders from the US president. Wolford (2017) also uses the Cox model to decribe the duration of peace among a coalition of victors in interstate wars.

Past Work

Freedman (2008) provides a critical primer to survival analysis from the perspective of medical research and experimentation. Box-Steffensmeier and DeBoef (2006); Box-Steffensmeier and Zorn (2001, 2002); and Box-Steffensmeier et al. (2003) introduce several of the more-complicated versions of the Cox model to political science audiences. Beck and Katz (2011) provide a broad reviews of the state of the art in the analysis of panel time series data.

Advanced Study

Therneau and Grambsch (2000) is an excellent resource for survival models. Kalbfleisch and Prentice (2002) present a detailed treatment as well. Box-Steffensmeier and Jones (2004) develop a political science-focused presentation of event history models. Park and Hendry (2015) argue for a revised practice in the evaluation of proportional hazards in the Cox model.

Time series analysis, much of which relies on the likelihood approach, is an area of vast research, with many texts available. The canonical text for analyzing time series and panel data remains Wooldridge (2010). Several recent contributions include Box-Steffensmeier et al. (2015); Brandt and Williams (2007); and Prado et al. (2017).

For an accessible introduction to spatial dependence in the context of regression models, see Ward and Gleditsch (2018). Beck et al. (2006) and Franzese and Hayes (2007) further discuss the interplay of spatial and temporal dependence.

Software Notes

Therneau and Grambsch (2000) underpins the `survival` library in $\mathcal{R}$. The `eha` package (Broström, 2012, 2017) provides other parameterizations of common survival models, including parametric models capable of handling time-varying covariates. `flexsurv` (Jackson, 2016) and `rms` (Harrell, Jr., 2017) are other alternatives that build on `survival`. The `spduration` library (Beger et al., 2017) implements the split-population survival model used in this chapter. See also `smcure` (Cai et al., 2012).