Appendix A:
Duration Model Overview

Duration models are concerned with lengths of time as dependent variable. They are useful when we are investigating questions about how long before a subject experiences some event. We count the number of periods the subject "survives" before experiencing the event; the resultant quantity is *t*, our duration of interest and dependent variable. OLS is not particularly well-suited to model durations, because it assumes (1) the dependent variable can be positive or negative, (2) that the regression equation's errors are normally distributed, (3) that all covariates are constant across the duration's length, and (4) that all our subjects fail during our observation period, meaning that we know the start and end point of every subject's duration (Crowder 2012, 3–5). One or more of these assumptions are false when working with durations as a dependent variable—quantities of time are non-negative, the errors they subsequently produce can be non-normal, covariates can be time varying, and we do not always observe the duration's end point for all our subjects (i.e., we may have right-censored durations).

Duration models come in three major classes. The classes are differentiated from one another based on two factors. The first factor is *baseline hazard functional form*. It refers to whether we assume a functional form connecting the baseline hazard (h_0) to our observed *t*'s, where the baseline hazard expresses the hazard of the duration's terminating event occurring when there are either no covariates in the model, or when all of the covariates are set to zero. The second factor is the covariate link function. Here, we refer to whether there are covariates in the model, because if there are, we must specify a function connecting these covariates to our observed *t*'s. We use *x* to refer to a single covariate, and *X* to refer to the entire set of included covariates.

We begin our discussion with continuous-time durations. The continuous vs. discrete distinction derives from the underlying process generating your event of interest, *not* the frequency with which you measure your data. Continuous-time durations come from an underlying process in which your event of interest could occur *at any instant,* similar to how hockey or soccer scoring could occur at any time when play is live.[22] If we assume no functional form for the hazard and have no covariates, we obtain a non-parametric duration model. The Kaplan-Meier curve is the classic example, and is usually expressed in terms of the survivor function *S*(*t*), which represents the fraction of subjects still at risk in *t*.[23] Its equation is (Klein and Moeschberger 2003, 92):

$$S(t) = \prod^{t < t_{\text{Max}}} \left(1 - \frac{d_t}{N_t}\right) \qquad 1$$

implying that the hazard function is:

$$h(t) = \frac{d_t}{N_t} \qquad 2$$

where t_{Max} represents the largest recorded failure time, d_t represents the number of subjects that fail in time period *t*, and N_t represents the number of subjects still at risk at the start of *t*, which some denote as N_{t-}.

If we continue to impose no functional form for the baseline hazard, but include covariates in our specification, we now have a semi-parametric duration model. Cox proportional hazard models are the most ubiquitous of the semi-parametric duration models. Cox models take the form:

$$h(t) = h_0(t)\exp(\beta' X) \qquad 3$$

Notice how the covariates (*X*) and their corresponding estimates (β) are linked to *h*(*t*) using an exponential link function—we are exponentiating the linear score of β and *X*'s product.

Finally, if we do make a functional form assumption about the baseline hazard, with or without covariates, we obtain a parametric duration model. There are many variants of parametric duration models. They differ from one another in terms of the baseline hazard's functional form. For example, a Weibull parametric duration model, when specified in terms of proportional hazards (denoted with "PH" subscript), has the form:

---

[22] By contrast for discrete-time durations, to continue the sports analogy: baseball and American football scoring plays are inherently from a process in which time is discrete. Baseball runs can only occur for every *attempted pitch* in a game, while American football scoring can only occur for every *attempted play* in which the ball becomes live.
[23] If a semi-parametric duration model does not include covariates, the resulting *h*(*t*) is a non-parametric expression.

$$h(t) = pt^{p-1}\exp(\beta_{PH}{}'X) \quad 4$$

where p is the hazard's shape parameter, a quantity to be estimated from the data. Box-Steffensmeier and Jones (2004, chap. 3) provide a list of parametric duration models commonly used in political science, along with the models' functional forms.

On a concluding note, duration models' estimated coefficients can come in one of two metrics: accelerated failure time (AFT) or proportional hazards (PH). The coefficients' metric matters because a positive AFT coefficient means the *opposite* of a positive PH coefficient. Some duration models' coefficients can be reported in either metric (e.g., Weibull, exponential),[24] but some models' coefficients can only be reported in AFT (e.g., generalized Gamma, log-logistic, log-normal) or only in PH (e.g., Gompertz, Cox). Whether or not a model can be reported in AFT or PH depends on whether the model's underlying hazard can be reexpressed in a certain way. Coefficients in an AFT metric speak directly to x's effect on the *duration*, where a positive β_{AFT} means higher levels of x will lengthen t, the duration of interest. By contrast, coefficients from PH metric speak in to x's effect on the *hazard* of the duration ending ($h(t)$). Specifically, the hazard acts like a conditional probability, modeling the probability that subject i will experience the event terminating our duration in time period j, given that i has not experienced the event yet.[25] A positive β_{PH} means higher levels of x increase the probability of the duration's terminating event, implying the duration itself will be shorter.

[24] Importantly, the AFT vs. PH metrics for Weibull and exponential parametric models are based on the same underlying math. You can transform AFT coefficients into PH coefficients or vice-versa (Box-Steffensmeier and Jones 2004, 29).

[25] Continuous-time hazards are technically conditional risks, not probabilities; see Appendix B, Section 3.

Appendix B:
Different Interpretation Techniques

Our theme throughout is that predicted durations are typically more straightforward to interpret than predicted hazards, though they too have limitations. However, some duration models cannot generate predicted durations easily, making hazard-based interpretations the only option. Hazards have few, if any, quantities that are easy to interpret. Transition probabilities fill this specific need, as they are computed using hazards, but are interpretable as probabilities.

1. EXPECTED DURATION/MEDIAN DURATION

One way to interpret duration models is to predict *t* from its estimates, akin to how we can predict *y* from OLS models. In OLS, we take expectations to yield:

$$E(y|X) = \alpha + \beta_1 x_1 + \beta_2 x_2$$

which is what we usually call "predicted *y*" ($\hat{y}$).

Duration models are more complicated, because some of the parametric models' underlying distributions are not symmetric. The Weibull parametric duration model is one example. Because of the non-symmetry, two quantities may be of interest to us: the mean duration time (via expectations) and the median duration time. The expression for each quantity varies depending on the baseline hazard's functional form (for formulas, see Box-Steffensmeier and Jones 2004, chap. 3).

This interpretation technique works well for any parametric duration model, regardless of metric. Deriving the mean and median *t* for semi-parametric Cox models is more difficult, because both formulae involve an assumption about the baseline hazard's functional form. The Cox's defining feature is that it does *not* assume anything about the baseline hazard's functional form. Kropko and Harden (forthcoming) provide a technique for recovering expected durations from a Cox model, by using a generalized additive model (GAM) to regress the Cox's predicted $\exp(\beta_{PH}{}'X)$ against *t*, the observed duration. For non-parametric models, the mean and median are also derivable expressions by way the Kaplan-Meier estimate of *S*(*t*) (see Klein and Moeschberger 2003, 32–33).

2. Marginal Effects/First Differences/Time Ratio

We can also calculate the effect of a change in x on the expected duration, if we so desire.[26] It is akin to interpreting β in OLS, because OLS β estimates are equivalent to marginal effects. If x is continuous, we can take the partial derivative of $E(t|X)$ with respect to x, yielding x's marginal effect. Marginal effects tell us how much $E(t|X)$ changes for an infinitesimally small increase in x's value from some value ($x = 0$, or 1, 2…). If x is non-continuous, we can take the discrete first difference. The first difference would tell us how much $E(t|X)$ changes for a one-unit increase in x from some value. First differences are also perfectly acceptable for a continuous x. Licht (2011, 230) nicely summarizes the differences between marginal effects and first differences, in the context of interpreting interaction effects in duration models. Her summary generalizes to models without interaction terms. In practice, we prefer calculating first differences, regardless of whether x is continuous or not. We find it easier to speak of specific changes in x's value, instead of a nebulous "infinitesimally small increase," and how those specific changes correspond to changes in the mean duration. However, we reiterate that both marginal effects and first differences are correct, from a technical perspective.

It is also possible to interpret x's effect using a time ratio. Time ratios express, for a one-unit increase in x, how many times larger or smaller the mean duration becomes, holding all else equal. Time ratios work for models in an AFT metric, and are equal to $\exp(\beta_{\text{AFT}})$. Box-Steffensmeier and Jones refer to these in terms of "relative expected duration time" (2004, 29). Time ratios' use is rare in political science, but we mention them because of the analogous quantity for hazard-based models (hazard ratios), which we discuss in Appendix B, Section 4. Our general remarks there apply to time ratios as well.

[26] If we wanted, we could also calculate x's marginal effect on the median duration, instead of $E(t|X)$ (Box-Steffensmeier and Jones 2004, 31). We use the mean duration throughout this paragraph for explication purposes.

3. HAZARD RATE

In continuous-time duration models, the hazard represents the instantaneous rate at which the termination event occurs. We can interpret x's effect with respect to the hazard of the event occurring. Hazard-related interpretations are common for duration models whose βs come only in a PH metric, with Cox models being the canonical example. The generic form for any (proportional) hazard is:

$$h(t) = h_0(t)\exp(\beta_{PH}{}'X) \quad 5$$

The first hazard-related interpretation involves predicting $h(t)$'s value for a particular covariate profile, similar to how we can predict t for some models. $h_0(t)$'s functional form will vary, depending on what type of duration model is estimated.

This interpretation technique works well for parametric duration models, because we impose a functional form for $h_0(t)$. The technique works less well for Cox models, since we do not impose a functional form for $h_0(t)$, but it is possible to recover an estimate for $h_0(t)$ after estimating the model. Some software packages provide a canned procedure for plotting a smoothed estimate of a Cox's $h(t)$, which automatically recovers $h_0(t)$ when creating the plot, as we discuss in Appendix D. In lieu of a canned procedure, the cumulative hazard, $H(t)$, is an easier quantity to plot for Cox models because it can be estimated from the survivor function: $H(t) = -\ln(S(t))$.

The major drawback for hazard rates is their interpretability. Hazards, as a predicted quantity, are arguably unique to duration models with little analogue elsewhere, which hampers scholars' ability to effectively communicate substantive findings to a broader scholarly, or general, audience.[27] At best, hazards represent a conditional probability (discrete-time durations): the probability that the duration's terminating event occurs, given that it has not occurred yet. At worst, they represent an instantaneous risk (continuous-time durations), with a similar interpretation as marginal effects: it is the risk of experiencing

[27] The notable exception is sample selection models. The selection equation generates the inverse of Mills' ratio (IMR), which is then included as a regressor in the outcome equation. The IMR is, in fact, a hazard (Heckman 1979).

the duration's terminating event for an infinitesimally small increase in t's value. Neither possibility is particularly appealing, nor straightforward to convey easily.[28]

3.1. Cumulative Hazard and Survivor

Researchers also use two other absolute quantities to interpret duration models, which are particularly prevalent techniques for interpreting semi-parametric models. The first is $H(t)$, the cumulative hazard, which we mentioned above. $H(t)$'s slope speaks to the rate at which $h(t)$ changes, and can be larger than 1. $H(t)$ can be useful for comparing various covariate profiles, because $H(t)$ graphs will show which covariate profile has the highest risk of the event occurring to date. However, $H(t)$'s biggest drawback is that its magnitude is not particularly informative on its own (Singer and Willett 2003, 488).[29] Researchers require additional context to get a sense of an event's risk of occurrence *overall*. For instance, a cumulative hazard equal to 1.2 for a particular covariate profile tells us nothing. We would need another covariate profile or two, to be able to get a sense of whether this value is large or small by making relative comparisons.

The second quantity is $S(t)$, the survivor function, defined formally as $S(t) = \Pr(T \geq t)$. The survivor function provides information about how many subjects' failure times (denoted T in the formula) are above a specific time (usually denoted t). For instance, if we had five subjects who each failed at t = 1, 2, 7, 8, and 10, $S(t = 5) = 0.6$, since three subjects have failure times larger than 5.[30] We like the survivor function, because it is a close relative of transition probabilities, a newer technique we discuss in a moment. There are highly specific situations where transition probabilities will actually equal $F(t) = 1 -$

[28] Hazard rates also have a count-based interpretation, arguably the least inaccessible of all the possible interpretations. It begins by allowing events to occur repeatedly, hypothetically. A hazard rate then represents the expected number of events that would occur in one time period, assuming the rate stays constant within that time period (Cleves et al. 2010, 15). Hazard rates' risk-based interpretation is more common.

[29] Cumulative hazards can have a similar count-based interpretation as $h(t)$ (fn. 28): it is the number of events we would (hypothetically) expect to see from $t = 0$ through the current time period.

[30] Sedgwick and Joekes (2013) provide some examples of interpreting the Kaplan-Meier, along with common misconceptions about what the quantity represents.

$S(t)$.[31] Teachman and Hayward (1993) showcase $S(t)$'s usefulness for interpreting hazard-based models, and provide example interpretations. To the extent $S(t)$ (or $F(t)$) have weaknesses, they are less useful for interpreting more complex duration models, such as competing risks and repeated events. However, they are not alone. Many of the other interpretation strategies we list also become more unwieldy to use for interpreting more complex duration models, with the exception of transition probabilities.

4. Hazard Ratio, Percent Change in Hazard Rate

For relative comparisons using the hazard, one option is the hazard ratio, also sometimes called the risk ratio or relative risk (Box-Steffensmeier and Jones 2004, 50; Mills 2011, 94). The hazard ratio is equal to $\exp(\beta_{PH})$, and it describes "the effect of a one-unit difference in the associated predictor on the…hazard" (Singer and Willett 2003, 524). It is a ratio because it compares $h(t)$ from one covariate profile ($x_{(1)} = 0$, say) to $h(t)$ from another covariate profile ($x_{(2)} = 1$), and expresses how to obtain $h(t|x_{(2)})$ as a multiple of $h(t|x_{(1)})$. Or, more plainly: if we think of covariate profile $x_{(1)}$'s hazard as an image, $\exp(\beta_{PH})$ represents how much you would have to shrink or enlarge that image to obtain covariate profile $x_{(2)}$'s hazard. This is evident from inserting $x_{(1)} = 0$ and $x_{(2)} = 1$ as arbitrary values into equation 5 and simplifying terms:

$$h(t|x_{(1)}) = h_0(t)\exp(\beta_{PH}x_{(1)}) \qquad h(t|x_{(2)}) = h_0(t)\exp(\beta_{PH}x_{(2)}) \tag{6}$$

$$\frac{h(t|x_{(2)})}{h(t|x_{(1)})} = \frac{\exp(\beta_{PH}x_{(2)})}{\exp(\beta_{PH}x_{(1)})}$$

$$\frac{h(t|x=1)}{h(t|x=0)} = \exp(\beta_{PH}[1-0])$$

[31] The two will be equal for a classic duration transition structure—all subjects start in stage 1, two possible stages, and the only transition possible is 1→2—when the transition probability interval begins at $s = 0$. In a similar vein, our transition probabilities will equal the cumulative incidence function (CIF) in a classic competing risks situation—all subjects begin in stage 1, there are any number of additional stages, and the only possible transitions are out of stage 1—again when the transition probability interval begins at $s = 0$ (Putter, Fiocco, and Geskus 2007, 2424).

$h(t|x=1) = \exp(\beta_{PH})h(t|x=0)$

This interpretation technique is appealing for Cox models, because the baseline hazard terms cancel out in the ratio, eliminating the complications arising from $h_0(t)$ in a semi-parametric expression. Because of the exponentiation, we are interested in seeing whether $\exp(\beta_{PH})$ is statistically different from 1 for hypothesis testing purposes, since exp(0) = 1. A hazard ratio greater than 1 means that a $(x_{(2)} - x_{(1)})$-unit change from $x_{(1)}$ produces a higher hazard rate. A hazard ratio less than 1 means that a $(x_{(2)} - x_{(1)})$-unit change from $x_{(1)}$ produces a lower hazard rate. The hazard ratio's major disadvantage is yet again its interpretability. Hazard ratios are not particularly intuitive quantities, perhaps because they refer to the hazard's scale across two different covariate profiles. This is especially true for audiences not already familiar with hazards, as we have already mentioned. Hazard ratios are also relative measures of change, meaning they may be misleading for infrequent events, as we discuss in the context of percent changes in $h(t)$.

Finally, we can compute *x*'s effect on $h(t)$ via the marginal effect or first difference. The generalities of our marginal effects/first difference discussion, when *t* was the quantity of interest, also apply here. For $h(t)$, the percent change in the hazard's value is similar to the first difference (though not equivalent), and is used more frequently. To calculate how some change in *x*'s value affects $h(t)$ in terms of percent change, the formula is (Box-Steffensmeier and Jones 2004, 60; Mills 2011, 95):

$$\%\Delta h(t|X) = \frac{\left(\exp(\beta_{PH}x_{(2)}) - \exp(\beta_{PH}x_{(1)})\right)}{\exp(\beta_{PH}x_{(1)})} * 100 \quad (7)$$

$$\%\Delta h(t|X) = \left(\exp(\beta_{PH}[x_{(2)} - x_{(1)}]) - 1\right) * 100$$

where $x_{(1)}$ and $x_{(2)}$ denote the *x* values of interest. Earlier, we used $x_{(1)} = 1$ and $x_{(2)} = 0$.

Box-Steffensmeier and Jones (2004, 60) showcase $\%\Delta h(t|X)$ as a preferred way for interpreting Cox models, for reasons that make sense. h(t) is not an intuitive quantity, but percentages at least appear in other contexts, providing some modicum of familiarity to scholars. $\%\Delta h(t|X)$ does nevertheless have limitations, aside from $h(t)$'s interpretability. In particular, as with any method focusing on percent changes in *y*'s value, the quantity may be misleading for relatively infrequent events. For instance, it may be that increasing *x*'s value produces a 100% increase in the hazard rate. However, in substantive terms, this may be a very small change in the likelihood of an event actually occurring. A 100% increase the hazard could come about if $h(t)$ increased in value from 0.4 to 0.8 (fairly frequent event), but it could also come about if the hazard increased from 0.00001 to 0.00002 (very infrequent event). We simply cannot tell from percentage changes alone.

5. Transition Probabilities

Transition probabilities are a newer interpretation strategy. Their origin is in the multi-state duration model literature (see Metzger and Jones 2016 for further details). We also write further about them elsewhere (Metzger and Jones 2019). Despite the strategy's newness, we still include it because we are intrigued by the quantity's longer-term potential, hence our continued work in this area. Transition probabilities interpret duration model results from a new angle,[32] making the models potentially appealing to scholars who find duration-based or hazard-based interpretations unappealing.

Transition probabilities follow from the insight that any duration application can equivalently be conceived of as a transition between two or more stages. For instance, in studies of democratization, we might be interested in studying when non-democracies become democracies, thus the event in question is democratization. This can be equivalently thought of as countries occupying one of two stages, either a non-democratic stage, or a democratic stage, and our interest is understanding when countries transition from the former to the latter. Under this stage conceptualization, we are interested in not only *when* a transition between stages occurs, but also, *which* of the two stages a country occupies at any given point in time. Transition probabilities aim to answer the latter question—what is the probability of being in one stage or another.

Transition probabilities express the probability of being in some stage *by* a particular point in

[32] See fn. 31 for the specific instances where transition probabilities will be equivalent with $F(t)$ (= $1 - S(t)$) and the CIF.

time, given some set of starting conditions. We provide several applied examples where we interpret the quantities in our joint work (2018; Metzger and Jones 2016, 2019) and other work (Daniel and Metzger 2018; Jones 2013; Mattiacci and Jones 2016). Researchers must specify three things to generate a transition probability: the stage a subject starts in, how long the subject has already been at risk, and a covariate profile.

Transition probabilities' major strength is the ease with which they can be interpreted, since they are probability-based. Further, they are computed from cumulative hazards, meaning they take one of the more hard-to-interpret duration model quantities and transform them into something more useful. Transition probabilities can be calculated from any duration model, and if the quantity is simulated, they can be generated from models with any number of stage configurations. As far as drawbacks, the probabilities' computation time can be lengthy, since the two packages for generating Cox transition probabilities (`mstate` in R, `mstatecox` in Stata) both employ simulations. This is particularly true when the transition probabilities' magnitude is low, because in order to get sufficiently precise quantities for inference making, you will need a large number of subjects for each simulation pull.

Most of the remaining drawbacks are all related to human error. Like any predicted quantity, transition probabilities are only as sound as the estimated model they are generated from. If the estimated model is wrong, the probabilities will be, too. Similarly, if you misunderstand how the probabilities are interpreted, then the conclusions you draw will likely be wrong. Transition probabilities tell you the probability of transitioning *by* a specific time, *not* the probability of transitioning *in* a specific time. To see the distinction: there is a big difference in saying a flight took off at 1pm and landed **by** 3pm, and saying that a flight took off at 1pm and landed **at** 3pm. The former is correct for transition probabilities, not the latter.

Appendix C:
Example Application, with Interpretations

To demonstrate the some of the interpretation strategies for duration models discussed in the paper, we rely on a simplified dataset of turnover in the European Parliament.[33] The unit of analysis in the dataset is the individual member of European Parliament (MEP), and the dependent variable is the number of terms the member serves before leaving office. We include 6 independent variables: whether the MEP's home country has a federal system, his/her age, whether the MEP holds a leadership position in the European Parliament, whether the MEP holds a committee leadership post, whether they belong to a far-left party, and a dummy variable for whether the MEP is female. All of our model estimates and predicted quantities in this appendix are computed in `R` 3.4.2. We report our semi-parametric Cox model containing these variables in Table 2. The first column presents the coefficient estimates from the model, and the second presents hazard ratios, which will be discussed below. We focus on interpreting the various quantities here; for more on general explanations of each quantity, see Appendix B.

TABLE 2. MEP Turnover

	Coefficients	Hazard Ratios
Federal System	-0.267***	0.766
	(0.051)	(0.694, 0.846)
Age	-0.011***	0.989
	(0.003)	(0.984, 0.994)
Leadership Position in EP	-0.455***	0.635
	(0.092)	(0.530, 0.760)
EP Committee Leadership	0.010	1.01
	(0.063)	(0.893, 1.14)
Far Left	0.117	1.12
	(0.090)	(0.942, 1.341)
Female	-0.249***	0.779
	(0.060)	(0.693, 0.877)
Partial Log-Likelihood	-11975.28	

† = $p \leq 0.10$, * = $p \leq 0.05$, ** = $p \leq 0.01$, two-tailed tests. Standard errors reported in parentheses beneath coefficient estimates, 95% confidence intervals presented beneath hazard ratios in parentheses.

[33] The full dataset is based on Daniel (2015).

A. Hazard Ratio/% Change in Hazard Ratio

Hazard ratios can be calculated simply by exponentiating the coefficients from a semi-parametric Cox model. In R and Stata, these untransformed coefficients are provided by default when estimating a Cox model. In the model specification presented above, the indicator for whether an MEP's home country has a federal system has a hazard ratio of 0.766. Hazard ratios are relative to 1, with values below 1 indicating higher values of the independent variable *reduce* the hazard of an event occurring, and values above 1 indicating that higher values of the independent variables *increase* the hazard of an event occurring. Hazard ratios' interpretation is based on a one-unit increase in the independent variable's value, similar to a classic OLS coefficient interpretation. Thus, a hazard ratio of 0.766 indicates that an MEP from a federal system (= 1) is less likely to leave office than an MEP from a non-federal system (= 0). Specifically, the hazard of a federal MEP leaving office is 0.766 the size of a non-federal MEP's hazard of leaving office, holding all else equal. Moreover, as the 95% confidence intervals do not include 1, we can conclude the hazard ratio is statistically significant at the 95% level.

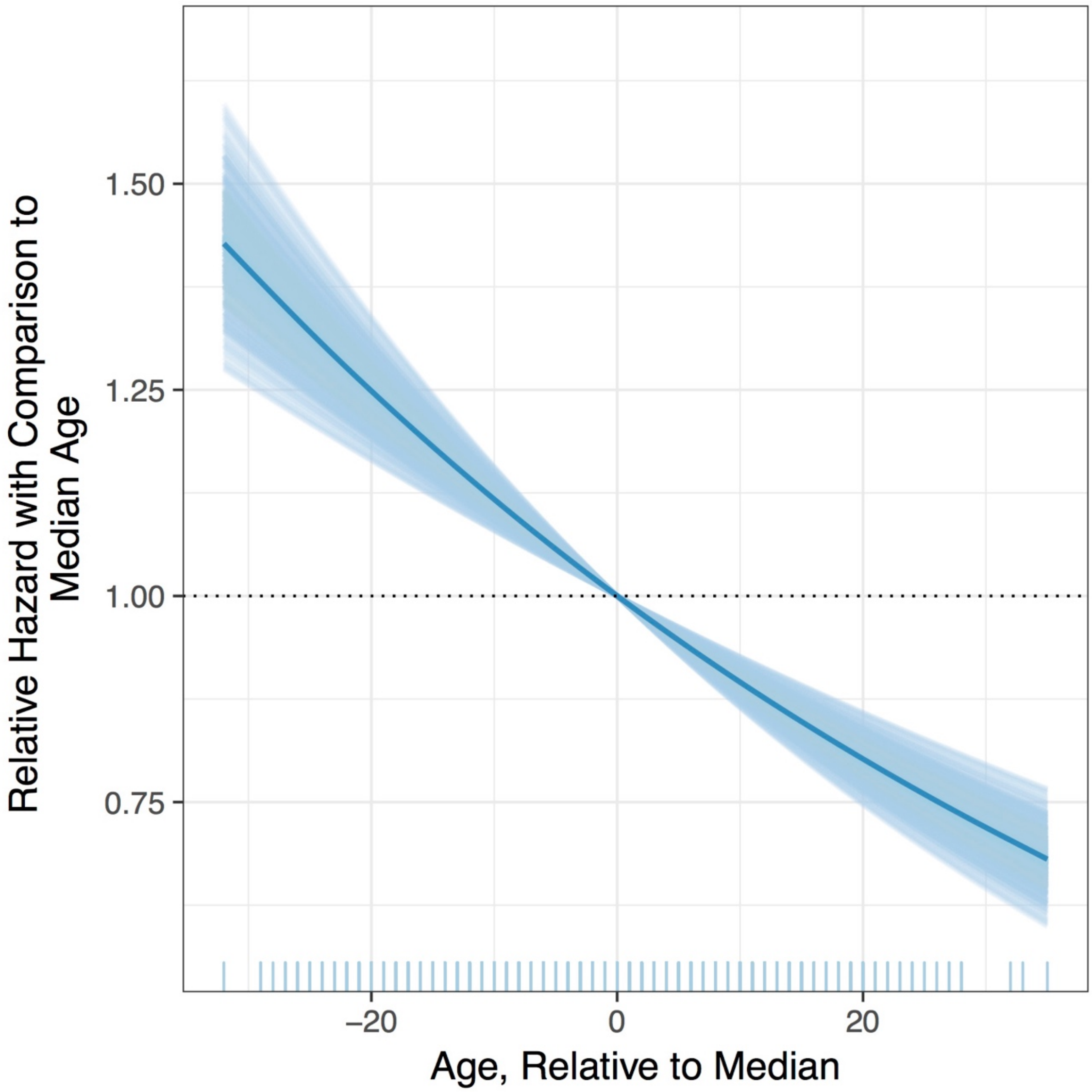
Hazard ratios can be easily transformed into a percentage scale by simply subtracting the hazard ratio from 1 and multiplying by 100. So, in the case above, a hazard ratio of 0.766 means if we were to change from a non-federal to a federal system (i.e., a change from 0 to 1), it would correspond with a (1 – 0.766) * 100 = 23.4% reduction in the risk of an MEP leaving office at the end of his/her term. The hazard ratio's statistical significance is typically calculated using the standard error for the estimated coefficient (β_{PH}). However, `R`'s `simPH` package also allows for simulated measures of uncertainty.

One drawback of relying on hazard ratios is that a one-unit change in the independent variable may not be meaningful in all circumstances. For example, the age variable ranges from 21 to 88. As a result, a one-unit change may not be especially informative. Similarly, if a variable is bounded between 0 and 1, a one-unit change would represent a change from the minimum to the maximum, which may not necessarily be the most reasonable comparison.

Instead, estimating relative hazards may be more instructive. Relative hazards offer a comparison between subjects with two fixed values that may be more or less than one unit apart. One useful strategy is to re-center the age variable around the median by simply subtracting its median value from all of the observed values. Doing so results in an age variable that is now centered around 0, with negative values indicating ages below the median, and positive values indicating ages above the median. With this re-scaling complete, we can now estimate relative hazards which compare scenarios in which age is 0 (the median) and other values, to achieve a more nuanced sense of how reasonable increases or decreases in an MEP's age alter the risk that s/he will experience turnover at the end of their term.

Figure 2 plots a sequence of these relative hazards ranging from the minimum to the maximum observed values of the re-scaled age variable. For each value along the *x*-axis, we estimate a relative hazard, comparing how a change from 0 (the median) to the value on the *x*-axis impacts the hazard of leaving office. `simPH` will simulate these values and automatically generate 95% confidence intervals (Figure 2's shaded region). As the figure shows, there is a clear relationship with values below the median significantly increasing the hazard of MEP turnover, whereas values above the median reduce the hazard of MEP turnover.

FIGURE 2. Relative Hazards of MEP Turnover



NOTE: Solid blue line represents median estimate, with other lines indicating simulated values. Dashed horizontal line denotes 1.

B. Survival Curves

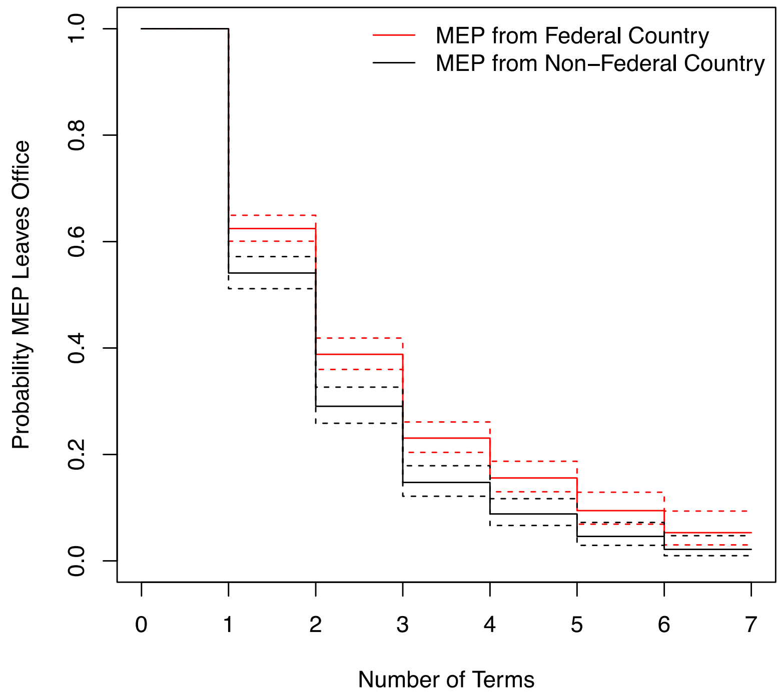
Survival curves are another common method for interpreting duration model results in political science. The survivor function, which is what we plot when graphing a survival curve, represents the probability that a subject will *not* experience an event in a particular year—i.e., the probability that the subject will continue to survive. In the context of MEP turnover, survival curves may offer some additional substantive appeal, as the question can be slightly reframed to: what is the probability of an

MEP remaining in office? When reframed in this way, survival curves offer a useful means for substantive interpretation.

Estimating survival curves and their corresponding confidence intervals is fairly straightforward with `R`'s `survival` package. Though survival curves have a more intuitive direct interpretation than some alternative strategies, such as hazard rates, it is nevertheless advisable to present two or more survival curves to offer a relative comparison of how changes in one or more covariates changes the probability of survival. We discuss why in the main text's Section II. Additionally, depending on the application, the probability of experiencing an event may be quite high, or quite low, so only presenting a single curve is potentially misleading.

With this in mind, we present two survival curves in Figure 3, varying whether the MEP's home country has a federal system or not, and holding all other independent variables at their median values. The probability of an MEP remaining in office—equivalent to *not* experiencing turnover—is higher when their home country has a federal system (red lines) than when it does not (black lines). Moreover, checking the confidence intervals' overlap indicates this effect is statistically significant following an MEP's first term, through their fourth term. However, following their fourth term in office, whether an MEP's home country has a federal system no longer has a statistically significant effect on their probability of remaining in office, as the confidence intervals for the two survival curves now overlap.

FIGURE 3. Survival Curves of MEP Turnover



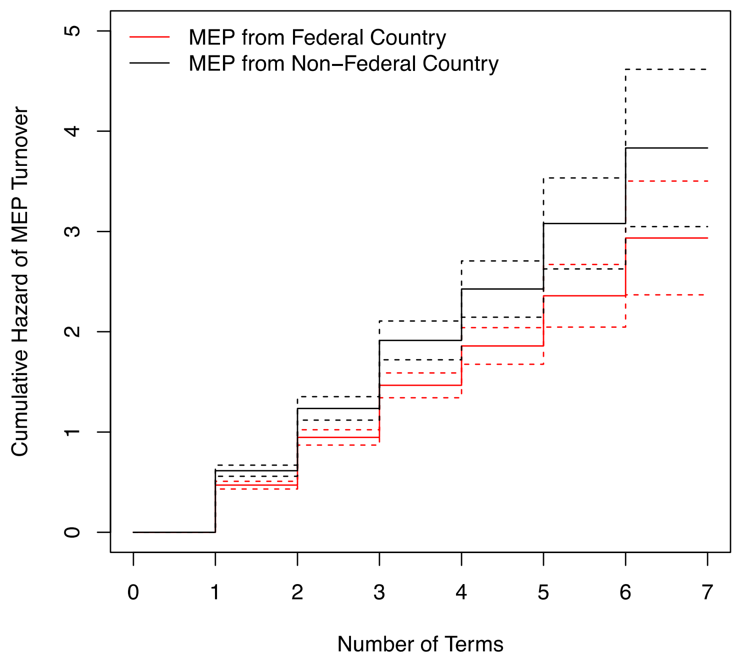
NOTE: Dashed lines represent 95% confidence intervals.

C. Cumulative Hazards

Cumulative hazards are related to survival curves, and are another potentially useful tool for substantive interpretation. The cumulative hazard function's value represents the overall risk of an event having occurred to date, though the values themselves are not directly interpretable beyond what we mention in the main text's fn. 20. Much like survival curves, cumulative hazards and confidence intervals are straightforward to estimate with `R`'s `survival` package.

We present two cumulative hazard estimates in Figure 4, varying whether an MEP's home country has a federal system or not, and holding all other covariates at their median values. As can be seen, MEP's from non-federal countries consistently have a higher risk of having experienced turnover (black lines) than their counterparts from federal countries (red lines). However, as was the case with the survival curves, this effect loses statistical significance at the 95% level after an MEP has been in office for five or more terms, indicated by the overlapping confidence intervals.

FIGURE 4. Cumulative Hazard of MEP Turnover

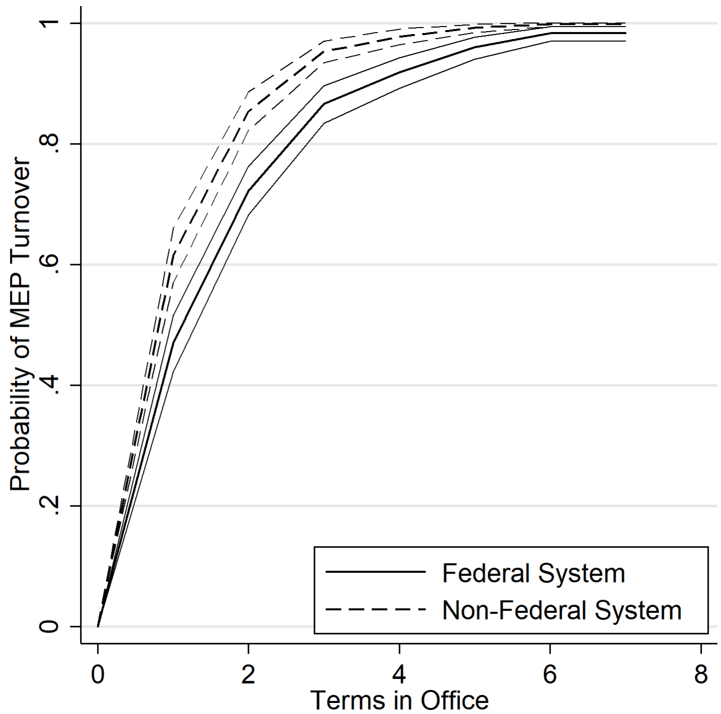


NOTE: Dashed lines represent 95% confidence intervals.

D. Transition Probabilities

To estimate transition probabilities, we need to begin by specifying three pieces of information: a covariate profile, in what stage subjects begin the analysis, and finally, what starting time the analysis begins in. For this application, we adopt naïve values for the latter two categories, specifying that the current stage is an MEP in office, and the starting time as 0, denoting that the MEP has just entered office for the first time. Finally, for covariate profiles, we vary whether the MEP's home country has a federal system or not, holding all other variables at their median values. Figure 5 depicts the simulated transition probabilities and 95% confidence intervals. As with each of the previous figures, we see MEPs from non-federal countries have a higher probability of leaving office than MEPs federal countries. Moreover, as with both the survival curves and cumulative hazards, this effect achieves statistical significance following an MEP's first term in office, and loses statistical significance after an MEP's fourth term.

FIGURE 5. Transition Probabilities of MEP Turnover



NOTE: Lines represent an MEP's probability of leaving office by *t*. All estimates begin with the current stage as MEP in office and *s* = 0. Quantities computed using simulation. Thin lines = 95% confidence intervals.

Appendix D:
Command List: Duration Model Quantities

Quick Reference Table

KEY
- = no
X = yes
Shaded cell = w/CIs or *p*-values
Slashed cell = N/A

	STATA			R		
Quantity	*NP*	*SP*	*P*	*NP*	*SP*	*P*
Duration-Based						
Absolute						
Mean/Expected Durat ($E(t)$)	X	X	X	X	X	X
Median Durat ($Q_{50}(t)$)	X	-	X	-	-	-
Relative						
Time Ratio	/	-	X	/	-	X
Marginal Effect	/	-	X	/	-	X
First Difference	/	-	X	/	X	X
Hazard-Based						
Absolute						
Hazard Rate ($h(t)$)	X	X	X	X	X	X
Relative						
Hazard Ratio	/	X	X	/	X	X
% Change in Hazard	/	X	X	/	X	-
Miscellaneous						
Survivor ($S(t)$)	X	X	X	X	X	X
Cumulative Hazard ($H(t)$)	X	X	X	X	X	X
Transition Probabilities	X	X	X	X	X	X

Note: this is not an exhaustive listing of duration-related commands, for either program.

Duration-Based Quantities (*t*)

- Absolute Quantities
  - Mean/Expected Duration
    - Stata
      - Non-Parametric
        - CAPABLE?: Yes, with built-in functionality.
        - COMMAND: `stci, rmean`
        - *P*-VALS/CIS?: Yes
        - OTHER REMARKS: Reports the restricted mean. Restricted mean also appears from `stdescribe` in "time at risk," "mean" cell. If no right-censored durations are present, restricted mean = true mean.
      - Semi-Parametric
        - CAPABLE?: Yes
        - COMMAND: `coxed` from Kropko and Harden (forthcoming)
        - *P*-VALS/CIS?: Yes
      - Parametric
        - CAPABLE?: Yes, with built-in functionality.
        - COMMAND: `margins, at(`*`COVAR VALUES`*`) predict(mean time)`
        - *P*-VALS/CIS?: Yes
        - OTHER REMARKS: Must first run `streg`.
    - R
      - Non-Parametric
        - CAPABLE?: Yes
        - COMMAND: Possible with the `survfit` function in the `survival` package. `print(survfit(Surv(t , fail) ~ 1), print.rmean = TRUE)`
        - *P*-VALS/CIS?: Yes
        - OTHER REMARKS: Also possible with `rms2` in `survRM2` package, which reports the restricted mean. Primarily designed for comparing mean survival times across two subgroups of the data.
      - Semi-Parametric
        - CAPABLE?: Yes
        - COMMAND: `coxed` from Kropko and Harden (forthcoming)
        - *P*-VALS/CIS?: Yes
      - Parametric
        - CAPABLE?: Yes
        - COMMAND: `mean_exp` in the `flexsurv` package after estimating a parametric model using `survreg` in the `survival` package.
        - *P*-VALS/CIS?: Yes
        - OTHER REMARKS: It is also possible to estimate a parametric model with `survreg` in the `survival` package, and then compute expected durations manually.

- Median Duration
  - Stata
    - Non-Parametric
      - CAPABLE?: Yes, with built-in functionality.
      - COMMAND: `stci, median`
      - *P*-VALS/CIS?: Yes
      - OTHER REMARKS: Also appears from `stsum` in the "Survival time," "50%" cell. If *S*(*t*) never drops below 0.5, the median is considered undefined. **Do *not* use `stdescribe` to obtain the median duration**. It can produce incorrect answers, because `stdescribe` calculates the median the way we would for any random variable. It finds the value of *t* for which half of the recorded *t* values are above and half below. Doing so *treats all the durations as observed* even if they are censored in truth, which is a problem. By contrast, the median duration is defined as the smallest value of *t* for which $S(t) \leq 0.5$, which *does* acknowledge right-censored durations may exist.
    - Semi-Parametric
      - CAPABLE?: None currently known.
      - OTHER REMARKS: Possible to calculate manually by visually examining Cox model's *S*(*t*) (see "Miscellaneous" > "Stata" > "*S*(*t*)" > "Semi-Parametric" in this appendix) and finding the smallest value of *t* for which $S(t) \leq 0.5$. If the model's *S*(*t*) never drops below 0.5, the median is considered undefined.
    - Parametric
      - CAPABLE?: Yes, with built-in functionality.
      - COMMAND: `margins, at(`*`COVAR VALUES`*`) predict(median time)`
      - *P*-VALS/CIS?: Yes
      - OTHER REMARKS: Must first run `streg`.
  - R
    - Non-Parametric
      - CAPABLE?: None currently known.
    - Semi-Parametric
      - CAPABLE?: None currently known.
      - OTHER REMARKS: Possible to calculate manually by visually examining Cox model's *S*(*t*) (see "Miscellaneous" > "R" > "*S*(*t*)" > "Semi-Parametric" in this appendix) and finding the smallest value of *t* for which $S(t) \leq 0.5$. If the model's *S*(*t*) never drops below 0.5, the median is considered undefined.
    - Parametric
      - CAPABLE?: None currently known.
      - OTHER REMARKS: It is possible to estimate a parametric model with `survreg` in the `survival` package, and then compute median durations manually.
- Relative Quantities
  - Time Ratio
    - Stata
      - Non-Parametric
        - CAPABLE?: N/A (no covariates in model, by definition)
      - Semi-Parametric
        - CAPABLE?: N/A; appropriate for models in accelerated failure time only.

- Parametric
  - CAPABLE?: Yes, with built-in functionality.
  - COMMAND: `streg varlist, dist(name) tr  // tr = key`
  - P-VALS/CIS?: Yes
  - OTHER REMARKS: Permissible option for AFT-capable parametric models only. p-values calculated from untransformed coefficients (β_{AFT}, vs. $\exp(\beta_{AFT})$) to avoid possible transformation-induced skew.[34]

- R
  - Non-Parametric
    - CAPABLE?: N/A (no covariates in model, by definition)
  - Semi-Parametric
    - CAPABLE?: N/A; appropriate for models in accelerated failure time only.
  - Parametric
    - CAPABLE?: Indirectly, yes
    - COMMAND: Possible via `aftreg` function in the `eha` package. `exp(coef(aftreg(function inputs here)))`
    - P-VALS/CIS?: Yes (same p-values as untransformed AFT coefficients)

- Marginal Effect
  - Stata
    - Non-Parametric
      - CAPABLE?: N/A. (No covariates to calculate marginals for, by definition.)
    - Semi-Parametric
      - CAPABLE?: None currently known. (Traditionally, appropriate for models in accelerated failure time only.)
    - Parametric
      - CAPABLE?: Yes, with built-in functionality.
      - COMMAND: `margins, dydx(x) at(VALUES) predict(QUANTITY)`
      - P-VALS/CIS?: Yes
      - OTHER REMARKS: Must first run `streg`. Appropriate for continuous covariates only. For discrete, use first differences. Permissible option for AFT-capable parametric models only. Quantities = `mean time` or `median time`
  - R
    - Non-Parametric
      - CAPABLE?: N/A. (No covariates to calculate marginals for, by definition.)
    - Semi-Parametric
      - CAPABLE?: Yes
      - COMMAND: `coxsimInteract` in the `simPH` package.
      - P-VALS/CIS?: Yes
      - OTHER REMARKS: Function primarily designed to estimate marginal effects for interaction terms in a model. The `simPH` package (Gandrud 2015) is a flexible package capable of generating numerous post-estimation quantities for Cox models. It generates CI estimates for all of its quantities of interest via simulation.

[34] See Stata 14 [ST] manual, p. 251.

- Parametric
  - CAPABLE?: Yes
  - COMMAND: `dydx` in the `margins` package.
  - *P*-VALS/CIS?: Yes
  - OTHER REMARKS: Function takes estimates from `survreg` in the `survival` package.

- First Difference
  - Stata
    - Non-Parametric
      - CAPABLE?: N/A. (No covariates to difference, by definition.)
    - Semi-Parametric
      - CAPABLE?: None known (Traditionally, appropriate for models in accelerated failure time only.)
      - OTHER REMARKS: Could indirectly calculate by generating the absolute-based quantity for multiple covariate profiles (see previous subsection), and then checking whether the profiles' confidence intervals overlap at a given *t*.
    - Parametric
      - CAPABLE?: Yes, with built-in functionality.
      - COMMAND: `margins, at(x=(v/v+1) z=VALUES) predict(QUANTITY) /// contrast(atcontrast(ar) pv)`
      - *P*-VALS/CIS?: Yes
      - OTHER REMARKS: Must first run `streg`. Permissible option for AFT-capable parametric models only. Quantities = `mean time` or `median time`
  - R
    - Non-Parametric
      - CAPABLE?: N/A. (No covariates to calculate marginals for, by definition.)
    - Semi-Parametric
      - CAPABLE?: Yes
      - COMMAND: `coxsimLinear` in the `simPH` package.
      - *P*-VALS/CIS?: Yes
    - Parametric
      - CAPABLE?: Yes
      - COMMAND: `dydx` in the `margins` package.
      - *P*-VALS/CIS?: Yes
      - OTHER REMARKS: Function takes estimates from `survreg` in the `survival` package.

<u>Hazard-Based Quantities (*h*(*t*))</u>

- Absolute Quantities
  - Hazard Rate
    - Stata
      - Non-Parametric
        - CAPABLE?: Yes, with built-in functionality.
        - COMMAND: `ltable _t _d, noadj haz`
        - *P*-VALS/CIS?: Yes
        - OTHER REMARKS: Can generate variable with `sts gen, h`, but will have no confidence intervals.

- Semi-Parametric
  - CAPABLE?: Indirectly, yes, through user-written package.
  - COMMAND: `mstsample, msfit` (`mstatecox` package)
    ```
    svmat tShoot_mstate, n(eqcol)
    // relevant quantities now in _time, _outhaz
    ```
  - *P*-VALS/CIS?: No
  - OTHER REMARKS: See Metzger and Jones (2018) for full instructions on using `mstatecox`. *h*(*t*) estimate from `mstsample` will be a step function, complements of the Cox's semi-parametric nature. (See, e.g., Box-Steffensmeier and Jones' Figure 4.1 (2004, 66)). Stata can plot a *smoothed* estimate of a Cox's *h*(*t*) via `stcurve` with no confidence intervals.
- Parametric
  - CAPABLE?: Yes, with built-in functionality.
  - COMMAND: `predict, haz`
  - *P*-VALS/CIS?: No
  - OTHER REMARKS: Must first estimate model with `streg`. Can set specific covariate values for the prediction by (temporarily) overwriting the observed covariate values with those from the scenario of interest. Can plot (without generating any variables) with `stcurve, haz`, but does not have the ability to plot confidence intervals.

▪ R

- Non-Parametric
  - CAPABLE?: Yes
  - COMMAND: `coxsimLinear` in the `simPH` package.
  - *P*-VALS/CIS?: Yes
- Semi-Parametric
  - CAPABLE?: Yes
  - COMMAND: `coxsimLinear` in the `simPH` package.
  - *P*-VALS/CIS?: Yes
- Parametric
  - CAPABLE?: Yes
  - COMMAND: `summary.flexsurv()` from the `flexsurv` package
  - *P*-VALS/CIS?: Yes
  - OTHER REMARKS: Use following the estimation of a parametric model using `survreg` in the `survival` package.

- Relative Quantities
  - Hazard Ratio
    - Stata
      - Non-Parametric
        - CAPABLE?: N/A (no covariates in model, by definition)
      - Semi-Parametric
        - CAPABLE?: Yes, with built-in functionality.
        - COMMAND: `stcox varlist, dist(name)  // no "nohr" option`
        - *P*-VALS/CIS?: Yes
        - OTHER REMARKS: p-values calculated from untransformed coefficients (β_{AFT}, vs. $\exp(\beta_{AFT})$) to avoid possible transformation-induced skew.[34]

- Parametric
  - CAPABLE?: Yes, with built-in functionality.
  - COMMAND: `streg varlist, dist(name)  // no "time" or "nohr"`
  - P-VALS/CIS?: Yes
  - OTHER REMARKS: Permissible option for PH-capable parametric models only. p-values calculated from untransformed coefficients (β_{AFT}, vs. $\exp(\beta_{AFT})$) to avoid possible transformation-induced skew.[34]

- R
  - Non-Parametric
    - CAPABLE?: N/A (no covariates in model, by definition)
  - Semi-Parametric
    - CAPABLE?: Yes
    - P-VALS/CIS?: Yes
    - COMMAND: `coxsimLinear` in the `simPH` package.
    - OTHER REMARKS: Also appears by default in `coxph` model object, as `exp(coef)` column.
  - Parametric
    - CAPABLE?: Yes
    - COMMAND: `WeibullReg` in the `SurvRegCensCov` package.
    - P-VALS/CIS?: Yes
    - OTHER REMARKS: This command employs the `survreg` function from the `survival` package, and then simply converts the results into alternate parameterizations.

- % Change in Hazard Rate
  - Stata
    - Non-Parametric
      - CAPABLE?: N/A. (No covariates to calculate marginals for, by definition.)
    - Semi-Parametric
      - CAPABLE?: Indirectly, yes, with built-in functionality.
      - COMMAND: `nlcom (exp(_b[x]*#)-1)*100`
      - P-VALS/CIS?: Yes
      - OTHER REMARKS: Must estimate `stcox` first. In code snippet, # = 1: increasing *x* by one unit; # = 2, increasing *x* by two units, etc.
    - Parametric
      - CAPABLE?: Indirectly, yes, with built-in functionality.
      - COMMAND: `nlcom (exp(_b[x]*#)-1)*100`
      - P-VALS/CIS?: Yes
      - OTHER REMARKS: Must estimate `streg` first. In code snippet, # = 1: increasing *x* by one unit; # = 2, increasing *x* by two units, etc.
  - R
    - Non-Parametric
      - CAPABLE?: N/A (No covariates to calculate marginals for, by definition)
    - Semi-Parametric
      - CAPABLE?: Yes
      - COMMAND: `coxsimLinear` in the `simPH` package.
      - P-VALS/CIS?: Yes
    - Parametric
      - CAPABLE?: None currently known.

Miscellaneous

- *S*(*t*)
  - Stata
    - Non-Parametric
      - CAPABLE?: Yes, with built-in functionality.
      - COMMAND: `sts gen newvar, s`
      - *P*-VALS/CIS?: Yes, with `sts gen newvar, lb(s)` and `sts gen newvar, ub(s)`
      - OTHER REMARKS: Can plot (without generating any variables) with `sts graph, surv ci`
    - Semi-Parametric
      - CAPABLE?: Yes, through user-written package.
      - COMMAND: `survci, surv` (`survci` package)
      - *P*-VALS/CIS?: Yes
      - OTHER REMARKS: Must first run `stcox`. See Cefalu (2011) for instructions on using `survci`. `survci`'s `outfile` option permits you to save multiple covariate profiles' *S*(*t*)s and their CIs to variables in a new dataset. Will only work with conventional standard errors (non-robust, no clusters). `stcurve` can also plot the survivor from a Cox, but does not have the ability to plot confidence intervals. For *S*(*t*) graphs after `tvc()`, albeit without confidence intervals, see Ruhe (2016).
    - Parametric
      - CAPABLE?: Indirectly, through user-written package.
      - COMMAND: `predictms, survival ci gen(varname)` (`multistate` package)
        ```
        // varname_1_1 = the survivor
        // varname_1_1_lci and _uci = confidence intervals
        ```
      - *P*-VALS/CIS?: Yes
      - OTHER REMARKS: See Crowther and Lambert (2017) for instructions on using `multistate` package. CIs come from simulations. Stata's built-in functionality is capable of either plotting *S*(*t*) (`stcurve, surv`) or saving *S*(*t*) to a variable (`predict, surv`), but neither generates CIs.
  - R
    - Non-Parametric
      - CAPABLE?: Yes
      - COMMAND: `survfit` in the `survival` package.
      - *P*-VALS/CIS?: Yes
    - Semi-Parametric
      - CAPABLE?: Yes
      - COMMAND: `survfit` in the `survival` package.
      - *P*-VALS/CIS?: Yes
      - OTHER REMARKS: Must first run `coxph`.
    - Parametric
      - CAPABLE?: Yes
      - COMMAND: Use `predict()` following the estimation of a parametric model using `survreg` in the `survival` package.
      - *P*-VALS/CIS?: Yes
      - OTHER REMARKS: The `predict()` function can be used to generate several quantities of interest following estimation of a parametric survival model including the linear predictor. Also possible with `summary.flexsurv()` from the `flexsurv` package following the estimation of a parametric model using `survreg` in the `survival` package

- *H*(*t*)
  - Stata
    - Non-Parametric
      - CAPABLE?: Yes, with built-in functionality.
      - COMMAND: `sts gen` *`newvar`*`, na`
      - *P*-VALS/CIS?: Yes, with `sts gen` *`newvar`*`, lb(na)` and `sts gen` *`newvar`*`, ub(na)`
      - OTHER REMARKS: Gives Nelson-Aalen estimate of cumulative hazard (the standard). Can plot (without generating any variables) with `sts graph, na ci`
    - Semi-Parametric
      - CAPABLE?: Yes, through user-written package.
      - COMMAND: `survci, cumh` (`survci` package)
      - *P*-VALS/CIS?: Yes
      - OTHER REMARKS: Must first run `stcox`. See Cefalu (2011) for instructions on using `survci`. `survci`'s `outfile` option permits you to save multiple covariate profiles' *H*(*t*)s and their CIs to variables in a new dataset. Will only work with conventional standard errors (non-robust, no clusters). `stcurve` can also plot the cumulative hazard from a Cox, but does not have the ability to plot confidence intervals.
    - Parametric
      - CAPABLE?: Yes, with built-in functionality.
      - COMMAND: `stcurve, cumh`
      - *P*-VALS/CIS?: No
      - OTHER REMARKS: Graphs only. Must first run `streg`. Can also obtain *H*(*t*) (without CIs) indirectly via *S*(*t*) (`predict, surv`), since *H*(*t*) = -ln(*S*(*t*)).
  - R
    - Non-Parametric
      - CAPABLE?: Yes
      - COMMAND: `survfit` in the `survival` package.
      - *P*-VALS/CIS?: Yes
    - Semi-Parametric
      - CAPABLE?: Yes
      - COMMAND: `survfit` in the `survival` package.
      - *P*-VALS/CIS?: Yes
    - Parametric
      - CAPABLE?: Yes
      - COMMAND: Use `summary.flexsurv()` from the `flexsurv` package following the estimation of a parametric model using `survreg` in the `survival` package.
      - *P*-VALS/CIS?: Yes
      - OTHER REMARKS: Confidence intervals generated via simulation.

- Transition Probabilities
  - Stata
    - Non-Parametric
      - CAPABLE?: Yes, through one of two user-written packages.
      - COMMAND: `predictms` (`multistate` package) --OR--
        `mstsample` (`mstatecox` package)
      - *P*-VALS/CIS?: Yes
      - OTHER REMARKS: See Crowther and Lambert (2017) for instructions on using `multistate` package; Metzger and Jones (2018) for full instructions on using `mstatecox`. CIs come from simulations for both packages.
    - Semi-Parametric
      - CAPABLE?: Yes, through user-written package.
      - COMMAND: `mstsample` (`mstatecox` package)
      - *P*-VALS/CIS?: Yes
      - OTHER REMARKS: See Metzger and Jones (2018) for full instructions on using `mstatecox`. CIs come from simulations.
    - Parametric
      - CAPABLE?: Yes, through user-written package.
      - COMMAND: `predictms` (`multistate` package)
      - *P*-VALS/CIS?: Yes
      - OTHER REMARKS: See Crowther and Lambert (2017) for instructions on using `multistate` package. CIs come from simulations.
  - R
    - Non-Parametric
      - CAPABLE?: Yes
      - COMMAND: `probtrans` following the use of `msfit` to generate cumulative hazards. Both commands located in the `mstate` package. The first step is to estimate a non-parametric model via `coxph` in the `survival` package.
      - *P*-VALS/CIS?: Yes
      - OTHER REMARKS: Measures of uncertainty can be generated either via analytical calculation, or through included bootstrap functionality, or simulation.
    - Semi-Parametric
      - CAPABLE?: Yes
      - COMMAND: `probtrans` following the use of `msfit` to generate cumulative hazards. Both commands located in the `mstate` package. The first step is to estimate a semi-parametric model via `coxph` in the `survival` package.
      - *P*-VALS/CIS?: Yes
      - OTHER REMARKS: Measures of uncertainty can be generated either via analytical calculation, the included bootstrap functionality, or simulation.

- Parametric
  - CAPABLE?: Yes
  - COMMAND: Use `msfit.flexsurv()` from the `flexsurv` package following the estimation of a parametric model using `survreg` in the `survival` package. This will generate cumulative hazards analogous to `msfit` in the `mstate` package. The resulting object is then compatible with `mstate`, and therefore, the `probtrans` command in `mstate` can be used to estimate transition probabilities from the original parametric estimates.
  - *P*-VALS/CIS?: Yes
  - OTHER REMARKS: As `probtrans` in `mstate` is being used to estimate transition probabilities, measures of uncertainty can be generated either via analytical calculation, through the included bootstrap functionality, or simulation.

Appendix E:
Toy Examples from Simulated Data

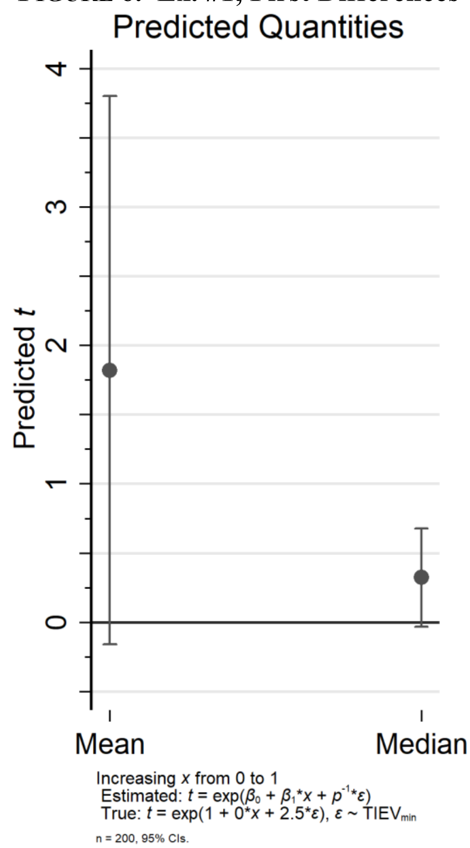
Situation #1: β vs. Predicted Quantity

TABLE 3. Ex. #1, Estimates

	Weibull
x	0.310**
	(0.147)
Constant	0.700***
	(0.168)
Shape (p)	0.561***
	(0.031)
N	200
N_{fail}	200
Log-Likelihood	-476.786

* = $p \leq 0.10$, ** = $p \leq 0.05$,
*** = $p \leq 0.01$, two-tailed tests.

FIGURE 6. Ex. #1, First Differences



Error distributed Type I Extreme Value (minimum), making Weibull the true DGP. Data generated in AFT. $N = 200$ and none of the data points are censored. $x \sim N(0,1)$.

While it may be rare for βs to show significance when there is truly no effect, it can happen. In our first toy example, we generate Weibull data where x's effect is 0, in truth. We then estimate a Weibull model with x included as a covariate (Table 3). At conventional levels of significance, the table alone suggests x's effect is statistically different from 0, with a two-tailed p-value of 0.036. However, when we generate predicted quantities using Table 3's estimates, a different picture emerges (Figure 6**Error! Reference source not found.**). We calculate the first difference in both the predicted mean and median when we increase x's value from 0 (its mean) to 1.[35] Neither first difference is statistically

[35] This one-unit change also happens to correspond to a one-standard deviation change in x's value, given how we generated x: $x \sim N(0,1)$.

different from zero, indicated by the 95% confidence intervals including zero (p-values = ~0.072). Had we only used the coefficients to assess our hypothesis, we would have drawn an incorrect conclusion.

Situation #2: Measures of Uncertainty

In our second toy example (Figure 7), we estimate a Cox semi-parametric duration model to assess whether x affects the event's occurrence and generate predicted quantities from this model.[36] From Figure 7's left panel, we would conclude increasing x's value produces an increase in the survivor function's value. Substantively, it implies higher x values make the event's occurrence less likely at a given time point, since higher values of $S(t)$ mean more subjects are still at risk of the event at time t. However, when we properly acknowledge the estimates' uncertainty, we see a change in x has no statistically significant effect (right panel). The two scenarios' 95% confidence intervals overlap.

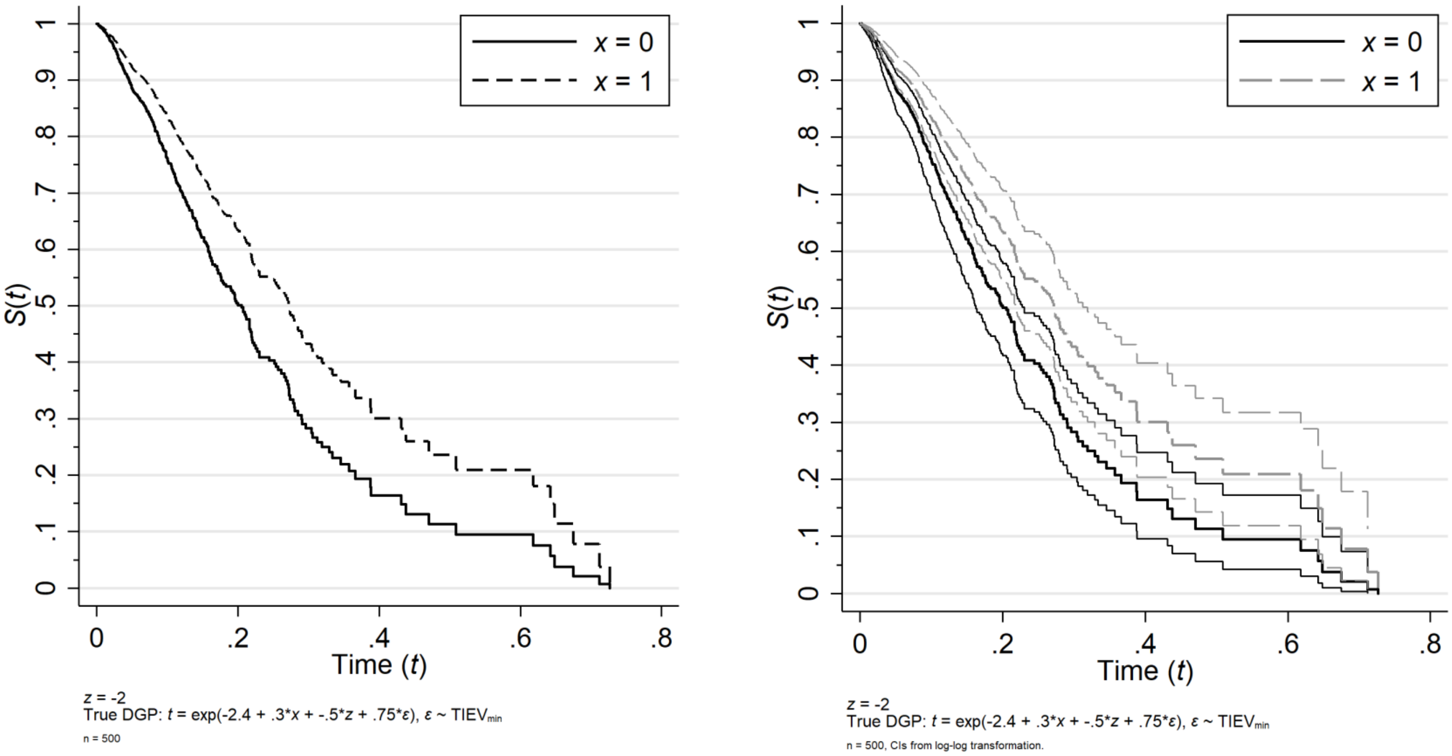
The phrase "statistically significant" is important. The Cox model estimates alone suggest x's effect is statistically different from zero. In truth, we know x does indeed have a non-zero effect on t because we generated these fake data. That non-zero effect, in turn, would produce a non-zero effect on $S(t)$'s values. We like this toy example because it highlights a related point: an effect may exist, but we must have enough data for our estimates to be sufficiently precise, given the effect's size and other sources of uncertainty in the model. Here, because of our sample's size (n = 500), we cannot conclude with confidence that the two predictions are statistically different. If we had data on more subjects, all else equal, the confidence intervals would be smaller.[37] Eventually, with a large enough sample the confidence intervals would not overlap at all, leading us to make (what we know to be, in this case) the 'correct' conclusion: x's effect is indeed non-zero.

[36] The true DGP is at the bottom of Figure 7's graphs.

[37] The same would be true if the model had less overall uncertainty, such as a smaller stochastic element. Practically speaking, though, researchers have little control over the amount of model uncertainty. Some uncertainty may be a product of the inherent process under study, if the process is stochastic; researchers cannot modify a true DGP in real life. In *relative* terms, by contrast, sample size is more within a researcher's realm of control.

FIGURE 7. Toy Example #2, *S*(*t*)

(a) No CIs (b) CIs



Error distributed Type I Extreme Value (minimum), making Weibull the true DGP. Data generated in AFT. $N = 500$, $x \sim N(0,1)$, $z \sim N(0,1)$, Corr(x,z) = 0.25, and none of the data points are censored.

Appendices Works Cited

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

Cefalu, Matthew. 2011. "Pointwise Confidence Intervals for the Covariate-Adjusted Survivor Function in the Cox Model." *Stata Journal* 11 (1): 64–81.

Cleves, Mario, William Gould, Roberto Gutierrez, and Yulia Marchenko. 2010. *An Introduction to Survival Analysis Using Stata*. 3rd ed. College Station, TX: Stata Press.

Crowder, Martin J. 2012. *Multivariate Survival Analysis and Competing Risks*. Boca Raton, FL: Chapman and Hall/CRC.

Crowther, Michael J., and Paul C. Lambert. 2017. "Parametric Multi-State Survival Models: Flexible Modelling Allowing Transition-Specific Distributions with Application to Estimating Clinically Useful Measures of Effect Differences." *Statistics in Medicine* 36 (29): 4719–42.

Daniel, William T. 2015. *Career Behaviour and the European Parliament: All Roads Lead Through Brussels?* Oxford: Oxford University Press.

Daniel, William T., and Shawna K. Metzger. 2018. "Within or Between Jobs? Determinants of Membership Volatility in the European Parliament, 1979–2014." *Journal of Legislative Studies* 24 (1): 90–108.

Gandrud, Christopher. 2015. "simPH: An R Package for Illustrating Estimates from Cox Proportional Hazard Models Including for Interactive and Nonlinear Effects." *Journal of Statistical Software* 65 (3).

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153–61.

Jones, Benjamin T. 2013. "The Past is Ever-Present: Civil War as a Dynamic Process." PhD diss. The Ohio State University.

Jones, Benjamin T., and Shawna K. Metzger. 2018. "Evaluating Conflict Dynamics: A Novel Empirical Approach to Stage Conceptions." *Journal of Conflict Resolution* 62 (4): 819–47.

Klein, John P., and Melvin L. Moeschberger. 2003. *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd ed. New York: Springer.

Kropko, Jonathan, and Jeffrey J. Harden. forthcoming. "Beyond the Hazard Ratio: Generating Expected Durations from the Cox Proportional Hazards Model." *British Journal of Political Science*.

Licht, Amanda A. 2011. "Change Comes with Time: Substantive Interpretation of Nonproportional Hazards in Event History Analysis." *Political Analysis* 19 (2): 227–43.

Mattiacci, Eleonora, and Benjamin T. Jones. 2016. "(Nuclear) Change of Plans: What Explains Nuclear Reversals?" *International Interactions* 42 (3): 530–58.

Metzger, Shawna K., and Benjamin T. Jones. 2016. "Surviving Phases: Introducing Multistate Survival Models." *Political Analysis* 24 (4): 457–77.

------. 2018. "mstatecox: A Package for Simulating Transition Probabilities from Semiparametric Multistate Survival Models." *Stata Journal* 18 (3): 533–63.

------. 2019. "Getting Time Right: Using Cox Models and Probabilities to Interpret Binary Panel Data." Working paper.

Mills, Melinda. 2011. *Introducing Survival and Event History Analysis*. Los Angeles: Sage.

Putter, Hein, Marta Fiocco, and Ronald B. Geskus. 2007. "Tutorial in Biostatistics: Competing Risks and Multi-State Models." *Statistics in Medicine* 26 (11): 2389–2430.

Ruhe, Constantin. 2016. "Estimating Survival Functions After stcox with Time-Varying Coefficients." *Stata Journal* 16 (4): 867–79.

Sedgwick, Philip, and Katherine Joekes. 2013. "Kaplan-Meier Survival Curves: Interpretation and Communication of Risk." *BMJ* 347: f7118.

Singer, Judith D., and John B. Willett. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford: Oxford University Press.

Teachman, Jay D., and Mark D. Hayward. 1993. "Interpreting Hazard Rate Models." *Sociological Methods & Research* 21 (3): 340–71.