Supplemental Appendices

Surviving Phases: Introducing Multi-state Survival Models

Shawna K. Metzger
National University of Singapore
smetzger@nus.edu.sg

Benjamin T. Jones
University of Mississippi
btjones1@olemiss.edu

Appendix A:
Multi-state Models: Technical Walkthrough

Multi-state models are organized around the concept of transitions—going *from* one stage *to* another stage. We use q to index the transitions. It is reasonable to think that the probability of a particular transition may vary across time—this is the very premise, after all, that motivates survival models.

The transition-specific hazard rate expresses the *instantaneous* risk of a particular transition occurring at a point in time. Mathematically, we can express this for a continuous-time duration as (Wreede, Fiocco, and Putter 2010, 262):[43]

$$\alpha_{gh}(t) = \lim_{\Delta t \to 0} \frac{\Pr(Z(t + \Delta t) = h | Z(t) = g)}{\Delta t} \qquad 5$$

The quantity to the left of the equals is the hazard rate (α) of transitioning from Stage g to Stage h at time t. Δt represents the increment we add to t, for the 'next' time point in the future. We make this increment infinitesimally small, to the point of it being zero, with the limit notation. The limit as Δt approaches zero is how we obtain the *instantaneous* risk of the transition occurring in t. Formally, the hazard is equal to the probability of being in Stage h in time ($t + \Delta t$) when the current stage in t is Stage g (the numerator), divided by the size of our increment (the denominator).[44] The above expression holds for discrete-time hazards, too, when $\Delta t = 1$ (Allison 2014, 19).

We know already that transitioning from some stage (e.g., g) to another stage (e.g., h) can be expressed in terms of q, the transition index notation. When we further write Equation 5's hazard in Cox semi-parametric form, we get Equation 2 from the main text.

The instantaneous rate of transition is relevant if we want to say something about the transition probability more generally. Multi-state models can compute the probability of transitioning from g to h within a specified time interval (s,t]. We can express this as:

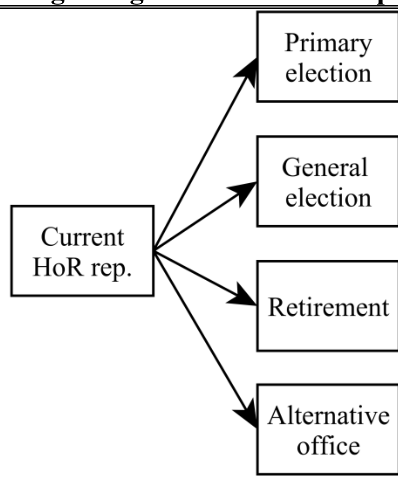
[43] This same general expression, albeit without transition-specific notation, undergirds basic survival models, too. See, e.g., Box-Steffensmeier and Jones (2004, 13).

[44] $Z(t)$ denotes the random process that determines the stage's value (e.g., Stage g, Stage h, etc.) in t.

$$\Pr_{gh}(s,t) = \Pr(Z(t) = h|Z(s) = g) \quad 6$$

The Markovian assumption is particularly visible in this expression (see Appendix C). The probability of occupying Stage h in time t is conditioned on the probability of being in g at the start of the time interval, s. We can—and do—split up the interval $(s,t]$ into smaller segments, which we denote with u. Formally, u is a list of all the discrete failure times within this interval. What this effectively means is: when any transition happens, the time at which the transition occurs is an element in u.

FIGURE 7. Stage Diagram – House of Representatives



It is easiest to show how multi-state's underlying math works with an example.[45] Let us say, for our US House example (Figure 7), we have (fake) data on 5 House representatives:[46]

[45] Beyersmann, Allignol, and Schumacher (2011, 177–182) also provide a reasonably accessible discussion of the calculation components, with the requisite mathematical notation.

[46] All five of our fake HoR reps experience a transition by $t = 3$; there is no right-censoring. This is for simplicity's sake. All of our calculations can accommodate handle right-censoring without a problem.

TABLE 3. Fake HoR Data

Subj. ID	t_0	t	Current Stage	Next Stage	Transition observed?
1	0	1	1 (In office)	2 (Primary el.)	1
2	0	1	1 (In office)	4 (Retires)	1
3	0	2	1 (In office)	3 (Gen el.)	1
4	0	3	1 (In office)	5 (Alt. office)	1
5	0	3	1 (In office)	4 (Retires)	1

We are interested in the transition probabilities for the interval (0,3], for all the stages. There are three u's in this interval: at 1, 2, and 3, because at least one transition occurs during each of these periods. To compute (non-parametric) transition probabilities, there are a few steps:

1. Calculate the transition-specific hazards for every discrete failure time point (u)
2. Calculate the transition probabilities by:
    a. Forming a matrix containing the hazards, for each u
    b. Adding together two matrices—an identity matrix and the hazard matrix from (a), for each u
    c. Multiplying every summed u matrix from (b) together

We begin by calculating the hazard for every transition (α_q). We need to do this for every period in which we observe *any* transition (denoted with u) between 1 and 3.[47] We know any hazard is equal to the number of failures we observe in a specific time period ("# Fails"), divided by the number of subjects at risk of experiencing failure at the start of the time period ("# at Risk") (Putter, Fiocco, and Geskus 2007, 2391–2392). For our example, the transition-specific hazards would be (light-shaded cells):

[47] Recall that our interval of interest is (0,3]. Notice how we exclude 0, the starting value (s), from our calculations, and begin with the next discrete time point, which is 1.

TABLE 4. Fake HoR: Hazards and Cumulative Hazards, by Transition

1 → 2					1→3				
u	# at Risk	# Fails	α_{12}	A_{12}	u	# at Risk	# Fails	α_{13}	A_{13}
1	5	1	0.20	0.20	1	5	0	0	0
2	3	0	0	0.20	2	3	1	0.33	0.33
3	2	0	0	0.20	3	2	0	0	0.33
1→4					1→5				
u	# at Risk	# Fails	α_{14}	A_{14}	u	# at Risk	# Fails	α_{15}	A_{15}
1	5	1	0.20	0.20	1	5	0	0	0
2	3	0	0	0.20	2	3	0	0	0
3	2	1	0.50	0.70	3	2	1	0.50	0.50

From here, we can calculate the transition probabilities. The formula is in Equation 4. We will need to add together two matrices for every u in our time interval: the identity matrix (**I**), and the hazard at every u ($\Delta\mathbf{A}$).[48] The two matrices' dimensions are determined by the number of stages we have, in a given application. Here, each matrix will be 5 x 5, since our HoR example has 5 stages. As we mentioned in fn. 16, which stage receives which number is irrelevant, so long as the numbering is consistent throughout the application.

We can start by writing the $\Delta\mathbf{A}$ matrix for $u = 1$. It will contain $u = 1$'s α_q values. The first row of this matrix represents all transitions *from* Stage 1. The columns, of which there are five, represent transitions *into* any other stage. The first value in the first row represents the 1→1 transition; the second, the 1→2 transition; the third, the 1→3 transition; and so on. Plugging in Table 4's α_q values, we get:[49]

[48] The "Δ" comes from the fact that, technically, we are interested in the change in the cumulative hazard's value (denoted A) between time period u and $u - 1$. The cumulative hazard is exactly what it sounds like—it accumulates (i.e., sums) the hazard for every time point within an interval. The cumulative hazard formula is Equation 3 in the main text; it is the Nelson-Aalen estimator. The dark-shaded columns in Table 4 contain the cumulative hazards for the fake HoR data. The change in the cumulative hazard's value between u and $u - 1$ is simply equal to the hazard in u; Table 4 illustrates this well.

[49] The second, third, fourth, and fifth rows in this matrix represent transitions from Stage 2, Stage 3, Stage 4, and Stage 5, respectively. All of these rows have zeros because we do not permit any transitions out of these stages in our HoR example (see Figure 7).

$$\begin{bmatrix} ? & 0.2 & 0 & 0.2 & 0 \\ 0 & ? & 0 & 0 & 0 \\ 0 & 0 & ? & 0 & 0 \\ 0 & 0 & 0 & ? & 0 \\ 0 & 0 & 0 & 0 & ? \end{bmatrix}$$

Row 1's diagonal's value represents the hazard of staying in Stage 1 (α_{11}). Intuitively, the hazard of staying is equal to 1 – (hazard of not staying). For α_{11}, this would be equal to $1 - \alpha_{12} - \alpha_{13} - \alpha_{14} - \alpha_{15}$. Ultimately, this is what we calculate with $\mathbf{I} + \Delta\mathbf{A}$; $\mathbf{I}$ represents "1" and $\Delta\mathbf{A}$ represents "– (hazard of not staying)". Since we are forming the $\Delta\mathbf{A}$ matrix, we must add up all of the *other* hazards (i.e., all the row's off-diagonal values), and multiply by -1 to get the diagonal's value.[50] In this case, row 1's off-diagonal values sum to 0.2 + 0 + 0.2 + 0 = 0.4. Row 1's diagonal is therefore 0.4 * -1 = -0.4. All the other rows sum to 0, making their diagonals also equal to 0.

Next, we need to sum the identity matrix and $\Delta\mathbf{A}$ matrix for every *u*. All of the relevant matrices, and their sums, look like:

u	$\mathbf{I} + \Delta\mathbf{A}$	Summed
1	$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} -0.4 & 0.2 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.6 & 0.2 & 0 & 0.2 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$
2	$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} -0.3 & 0 & 0.3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.7 & 0 & 0.3 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$
3	$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} -1 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

The final step is to multiply all the relevant *u* matrices (shaded column) together, indicated by the product integral in Equation 4. The resultant matrix is the full transition probability matrix (Equation 7),

[50] Formally, the $\Delta\mathbf{A}$ diagonals are equal to $-\Sigma_{h \neq g} A_{gh}(u)$ (Wreede, Fiocco, and Putter 2010, 262).

denoting the probability of transitioning from one stage to another, in the time interval 0 (exclusive) to 3 (inclusive).

$$P(0,3) = \begin{bmatrix} 0.6 & 0.2 & 0 & 0.2 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 0.7 & 0 & 0.3 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\boxed{P(0,3) = \begin{bmatrix} 0 & 0.2 & 0.2 & 0.4 & 0.2 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}} \qquad 7$$

For an example interpretation: cell 1,4 represents the probability of occupying Stage 1 (in office) in time period 0 and Stage 4 (retirement) in time period 3. We see there is a 40% chance that a House incumbent will be retired by time period 3.[51] The quantity demonstrates quite nicely how the transition probabilities work, and the proper interpretation of them. From the long-hand calculations, we know one subject transitions into retirement in time period 1, and another transitions in time period 3. By time period 3, both subjects are in the Retirement stage—that is, they are observationally equivalent. However, the math acknowledges the subjects' different transition times when we multiplied all the **I** + Δ**A** matrices together. For more complicated stage structures, the broader implication is that the multiplying takes care of any direct *and* indirect transitions from the starting stage to the ending stage of interest.

In another helpful illustration, notice how row 1 is the only row with off-diagonal quantities. This is because we only permitted transitions out of Stage 1 in our HoR example (see Figure 7). This is also why all of the non-row 1 diagonals are equal to 1. If subjects (somehow) occupy Stage 2 in time period 0, there is a 100% chance they will be in Stage 2 in time period 3 (cell 2,2), because there is no possible transition out of Stage 2.

[51] This quantity is not particularly surprising. Of the 5 subjects in our fake dataset, two of them retire in the specified interval; 2/5 = 0.4. The purpose of this example, though, is to show how the math behind multi-state models works.

Finally, row 1's diagonal also makes sense. In our fake dataset, all five subjects experienced a transition by the end of time period 3 (see fn. 48). As a result, there is a 0% chance that a subject would be in Stage 1 at period 0, and would still be in Stage 1 at period 3. This is the quantity in cell 1,1 of the transition matrix.

Appendix B:
Stages Rather than Events

Event history models are typically interpreted as modeling the time until an event occurs—for instance, how long until a war ends, a candidate exits a primary, or a government falls. However, an equivalent interpretation of event history models is that they, in fact, model the time that a subject spends within a particular stage. For instance, if the researcher is interested in modeling how long a war lasts, typically the question is framed as modeling the time until the end of the fighting is observed. Yet this same analysis can be equally understood as modeling how long a country remains in the "war" stage, with the country remaining within that stage until observing an event (the end of fighting). Importantly, these two interpretations are interchangeable. When the researcher's interest is modeling a single event, conceiving of the analysis as modeling the time until the event is observed is typically more straightforward. However, when the researcher is interested in multiple stages, as is often the case when modeling a process as a whole, the stage interpretation of event history models is often much more useful.

In this context, a stage simply denotes what happens to a subject in the time between two observed events (transitions). If we return to the war example, we may be interested not only in how long the war lasts, but also in how long peace will endure following the war. Thus, we can conceive of two broad stages: war and peace. If the analysis begins on the first day of a war, then a country will remain in the war stage until it experiences an observed event—the end of fighting. Once this event is observed, the country will transition to the subsequent peace stage. Following such an event, the country will remain in the peace stage until it experiences a subsequent event, for example, a return to fighting, which would convey a transition back to war. Therefore, stages simply reflect what happens to a subject after it experiences a particular event.

Appendix C:
The Markov Assumption

Multi-state models operate under a Markov assumption for their hazards. In econometrics, a (temporal) process is Markovian if the past and future are independent of one another, once we condition on the present (Hougaard 2000, 142). Or, more plainly: what comes next only depends on what is happening now, and nothing else. The Markov assumption hinges on the fact that the outcome we observe in the present (call this y_t) captures information about the process' entire history, up through t. When we predict the next period's outcome (y_{t+1}), controlling for y_t accounts for this history. This property is useful, econometrically, because we know omitted variables can bias our estimates if they are correlated with (included) covariates of interest. "History" might be one such omitted variable. Adding a control for t when predicting $t + 1$ (i.e., a one-period lag) can guard against this possibility.[52] Unsurprisingly, lags are common when working with longitudinal and time-series data, and much has been written about their use in these situations.

Multi-state models' hazards are Markovian in *stages*. They assume that a subject's next stage is conditional only on the subject's current stage, implying a subject's current stage effectively encapsulates its entire history. The Markovian stage-conditionality is evident in our generic hazard in Appendix A (Equation 5). In the numerator, we see the hazard of transitioning from Stage g to Stage h in time t is *only* dependent upon a subject currently being in Stage g; there is nothing else about a subject's history to the right of the conditional.[53] An important implication is that the transition-specific hazard is independent of how long a subject has been in the current stage (Hougaard 2000, 143).

Multi-state models' Markovian assumption is handy because it eases our computational burden (Hougaard 2000, chap. 6). Further, when the underlying process is indeed Markovian, modeling it as such will produce estimates with a smaller variance than alternative methods' estimates (Meira-Machado,

[52] Lagged dependent variables can introduce other econometric concerns, which others have explicated at length (Keele and Kelly 2006; Lebo and Moore 2003; Wawro 2002).

[53] The conditionality is also evident in how we define our transition probabilities (Equation 6).

Uña-Álvarez, and Cadarso-Suárez 2006). The Markovian assumption will hold, from a modeling perspective, when we measure *t* as elapsed time.

Of course, there may be instances in which we want to relax the Markovian assumption. History sometimes matters in political science, in substantively relevant ways. We may wish to explore these possibilities. When we relax the Markov assumption, we are taking a semi-Markovian approach.[54] We can formally tell when this is so based on the transition-specific hazards' specification (Hougaard 2000, 178). Markovian multi-state models have transition-specific hazards whose values do "not depend on any other aspect of the history, like states visited on the way, and the times of previous transitions (except to the extent that this information is reflected by the present state)" (Hougaard 2000, 143). If we have a hazard that does not meet these criteria, it is semi-Markovian in nature.

Notable semi-Markovian cases include:

1. When we measure time using a gap-time (clock-reset) approach (Putter, Fiocco, and Geskus 2007, 2415): when a transition occurs, *t* resets to 0, and begins counting anew. The transition-specific hazards are now modeling how long the subject will remain in the current stage. As a consequence, the hazards now depend on history, beyond what the current stage indicates, since *t* resets *any* time a transition occurs.
2. When we include a covariate for the time at which a subject has transitioned into a stage (Andersen and Pohar Perme 2013, 420; Putter, Fiocco, and Geskus 2007, 2416). These are also called "Markov extension models" when *t* is measured as total time (Hougaard 2000, 168–169). When *t* is gap time, the model is outright semi-Markovian.
3. When we include a covariate for the length of time spent in the previous stage.
4. When we include a set of covariates for previous stage (Beyersmann, Allignol, and Schumacher 2011, 198, 223–224). This can include a set of dummy covariates, or a count of previous visits to some other stage. Some also call these Markov extension models, rather than semi-Markovian (Mills 2011, 204).

[54] Semi-Markov models are also called "Markov renewal models", in some applications.

Appendix D:
Svolik 2015's Transition Frequencies
(Used to Replicate Maeda 2010)

TABLE 5. Observed Transition Frequencies – Svolik 2015

Current Stage	*Next Stage* Democracy	Exog.	Endog.	TOTAL
Democracy	–	49	24	73
(Row Total %)		(27.5%)	(13.5%)	(41.0%)
Exogenous	35	–	–	35
(Row Total %)	(83.3%)			(83.3%)
Endogenous	15	–	–	15
(Row Total %)	(68.2%)			(68.2%)
TOTAL	50	49	24	123

"TOTAL" column contains the row's total number of transitions, representing the number of transitions out of a stage. The percentage calculations use the number of countries in the stage as the denominator. They do not sum to 100% because of right censoring. Dashes indicate impossible transitions in our model.

Appendix E:
Transition Specific Covariates

In addition to testing for whether two or more transitions should be estimated as separate strata, we can also use our simulation analysis to draw attention to another valuable property of multi-state models: the ability to allow covariate effects to vary across transitions. Using the same simulation analysis described in the main text, we can assess whether our results will be biased if we fail to account for transition-specific covariate effects *out of* a stage, and if we fail to account for transition-specific covariate effects *into* a stage.

I. Exiting Transitions

In the context of straightforward competing risks models, this property is already well known to political scientists. Referring to Figure 1b, this would be analogous to a covariate, x, slightly increasing the risk of a subject transitioning from Stage 1 to Stage 2, but substantially increasing the risk that a subject will transition from Stage 1 to Stage 3. This is the scenario in our simulations below: we specify x will slightly increase the risk of transitions from Stage 1 into Stage 2 ($\beta_{(1\rightarrow2)x} = 0.5$), and will substantially increase the risk of transitions from Stage 1 into Stage 3 ($\beta_{(1\rightarrow3)x} = 3$). Table 6's left columns identify the true parameter values for each covariate effect (see Appendix I for all parameter values). We estimate a model in which we estimate a single coefficient for x's effect on transitions out of Stage 1. Or, more simply: we hold x's effect on the risk of transitioning out of Stage 1 constant and perform our simulations.[55]

TABLE 6. Simulation Results: Exiting Transitions

True β_x		*Estimated β_x (collapsed)*	
Parameter	*Value*	*Parameter*	*Value (Mean)*
$\beta_{(1\rightarrow2)x}$	0.5	$\beta_{(1\rightarrow2)x}$	1.262
$\beta_{(1\rightarrow3)x}$	3.0	$\beta_{(1\rightarrow3)x}$	1.262
$\beta_{(2\rightarrow1)x}$	1.0	$\beta_{(2\rightarrow1)x}$	0.759
$\beta_{(2\rightarrow3)x}$	-1.0	$\beta_{(2\rightarrow3)x}$	-0.902

Shaded cells are constrained to be equal during estimation.

[55] We estimate separate baseline hazards for Stage 1's two exiting transitions in our simulation.

We report our simulation results in Table 6's last column. As expected, the estimated coefficient for the collapsed transitions is biased (shaded rows), because *x* has a different effect on the probability of each transition. The estimated effect is approximately the average of *x*'s true effect on each transition.

II. Entering Transitions

Multi-state models can readily accommodate the above situation—competing risks models are, after all, a special case of multi-state models. However, a second, and more novel, scenario may also be of interest to researchers. It may be the case that the same covariate, *x*, exerts a different effect on the likelihood of a subject transitioning *into* the same stage, depending on the subject's current stage. The illness-death model, depicted in Figure 3 in the main text, is an example of precisely this type of scenario. *x* may have a different effect on the risk of a subject transitioning into Stage 3 depending on whether the subject is currently in Stage 1 or Stage 2.

Returning to our simulations, we specify that a covariate, *x*, will increase the risk of transitions into Stage 3 when the subject is in Stage 1 ($\beta_{(1\rightarrow3)x} = 3$), and will reduce the risk of transitions into Stage 3 when the subject is in Stage 2 ($\beta_{(2\rightarrow3)x} = -1$). Table 7's left columns identify the true parameter values for each covariate effect (again, see Appendix I for all parameter values). We estimate a model in which we estimate a single coefficient for *x*'s effect on transitions into Stage 3.[56] In other words, we hold *x*'s effect on the risk of transitioning into Stage 3 constant across transitions from Stage 1 and Stage 2, and perform our simulations.

TABLE 7. Simulation Results: Entering Transitions

True β_x		*Estimated β_x (collapsed)*	
Parameter	*Value*	*Parameter*	*Value (Mean)*
$\beta_{(1\rightarrow2)x}$	0.5	$\beta_{(1\rightarrow2)x}$	0.588
$\beta_{(1\rightarrow3)x}$	3.0	$\beta_{(1\rightarrow3)x}$	0.625
$\beta_{(2\rightarrow1)x}$	1.0	$\beta_{(2\rightarrow1)x}$	0.759
$\beta_{(2\rightarrow3)x}$	-1.0	$\beta_{(2\rightarrow3)x}$	0.625

Shaded cells are constrained to be equal during estimation.

[56] We also collapse both of the baselines hazards into Stage 3 (α_{13_0} and α_{23_0}) and include a time-varying indicator of whether a subject is in Stage 2; the two baseline hazards are equivalent, based on the parameter values we chose.

The results of this simulation are presented in Table 7's right columns. The results indicate that $\beta_{(1\rightarrow2)x}$ and $\beta_{(2\rightarrow1)x}$ remain unbiased. However, by failing to take into account the transition-specific effect of x on the risk of transitioning into Stage 3, the estimates of both $\beta_{(1\rightarrow3)x}$ and $\beta_{(2\rightarrow3)x}$ are now biased. Whereas the true effect of x on 1→3 is large and positive, by pooling the effect of x across transitions, the pooled coefficient is much smaller, understating the true effect of x on the risk of 1→3. Moreover, while the true effect of x on transitioning from 2→3 is negative, by pooling the effect of x on transitions into Stage 3, the coefficient that we recover is now positive, producing an erroneous finding with respect to the true effect of x.

Our two straightforward examples point to an important conclusion: researchers ought to be concerned with both identifying different covariate effects for different outcomes (i.e., exiting transitions from a stage) and also potentially different covariate effects for the same outcome (i.e., entering transitions), in the context of stages.

Appendix F:
Regime Switching Models

Our discussion in Section I.C makes clear how multi-state models are an example of a regime switching model. The phrase "regime switching model" is an umbrella term for a large class of models with many variations. Their common, defining characteristic is that they "allow the behavior of y_t [i.e., the DGP] to depend on the state of the system [S_t; i.e., the stage]" (Enders 2009, 439). Generically, for some y_t whose DGP has k covariates, regime switching models take the form:

$$y_t = \beta_{0_{S_t}} + \beta_{1_{S_t}} x_1 + \beta_{2_{S_t}} x_2 + \cdots + \beta_{k_{S_t}} x_k \qquad 8$$

where $S = \{1, 2, \ldots r\}$ is an index for each possible stage. S_t denotes the current stage in t. Notice how the estimates are subscripted with S_t, to indicate their values are dependent on the stage in t. Equation 8 has the same general form as multi-state models (Equation 2). There, the estimates are subscripted with q, the identifier for transitions, which permits a covariate's effects to vary based on the transition in question.

Multi-state models fill an arguable lacuna in the regime-switching literature.[57] Multi-state models pertain to durations, a quantity less talked about in the regime-switching context.[58] Additionally, in our multi-state model, S is known and observed by the researcher. Jackson (2011) discusses possible multi-state extensions for applications in which a process' stages are unobserved, using parametric survival models. We discuss these models further in Appendix J.III. Similarly, Spirling (2007, 396–399) sketches out a Bayesian approach for estimating a basic two-stage duration model, in which the stages are also unobserved. He, too, takes a parametric approach. We use semi-parametric survival models.

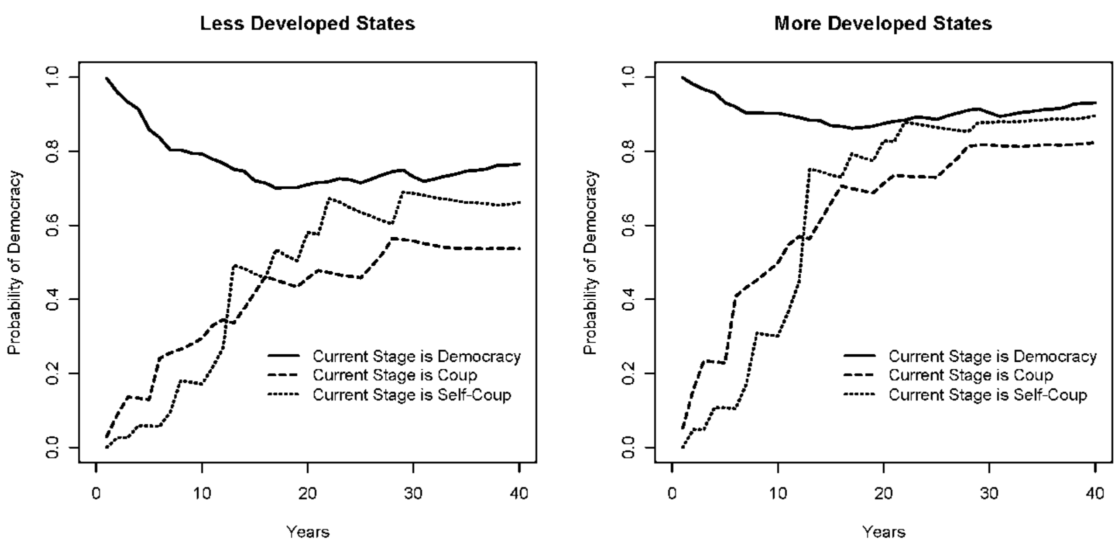
[57] For some general regime switching overviews, see Maddala and Kim (1998, chap. 15), Piger (2011), and Potter (1999).

[58] McGrath (2015) is a recent exception, in the context of two-stage discrete-time duration models. He points out that an event's onset and how long the event lasts are distinct stages, but practitioners often (unwittingly) pool these two stages together. Biased βs result, and this bias is even more complex in the presence of transition-specific covariate effects (2015, 537–538).

Appendix G:
Transition Probabilities: Varying the Starting Stage

Figure 8 extends the main text's results by highlighting an additional feature of transition probabilities: the ability to vary the stage a subject occupies at the beginning of the estimation. Figure 8 displays the probability that a country (1) stays democratic (D→D, solid line), (2) will be democratic after experiencing a coup (Ex→D, dashed), or (3) will be democratic after experiencing a self-coup (En→D, dotted), all by year *t*. The probabilities are calculated for $s = 0$, with all other covariates held at their median values. The probability of a country remaining non-democratic is quite low, especially in the case of more developed states. Figure 9 displays the same quantities, but adds confidence intervals.

FIGURE 8. Transition Probabilities into Democracy



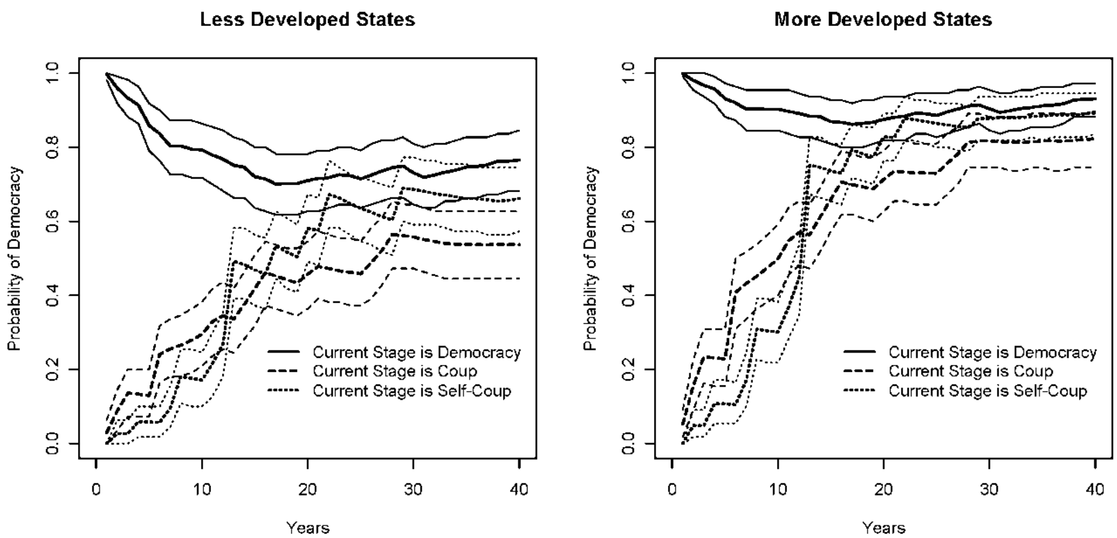
NOTE: Estimates begin at $s = 0$, reflecting the first year a country occupies a given stage. Each line denotes a different current stage. Quantities computed using simulation.

For set of example interpretations, using Figure 8's left-hand side:

- There is a 79.3% chance that a less-developed country will be in the Democracy stage at the 10-year mark, given it was a newly minted democracy (Democracy stage, $s = 0$).
- There is a 29.5% chance that a less-developed country will be in the Democracy stage at the 10-year mark, given it was a newly minted non-democracy via coup (Exogenous stage, $s = 0$).

- There is a 17.2% chance that a less-developed country will be in the Democracy stage at the 10-year mark, given it was a newly minted non-democracy via self-coup (Endogenous stage, $s = 0$).

FIGURE 9. Figure 8, with Confidence Intervals



NOTE: Estimates begin at $s = 0$, reflecting the first year a country occupies a given stage. Each line denotes a different current stage. Quantities computed using simulation. Thin lines = 95% confidence intervals.

Appendix H:
Simulated Transition Probabilities

We require simulation to compute transition probabilities for semi-Markov multi-state models (see Appendix C). We can also generate simulated transition probabilities for Markov multi-state models, if we desire. Generating simulated transition probabilities is a two-step process. We use our Maeda application as an illustration (Figure 4).

We begin by using `mstate`'s `mssample` function. `mssample` generates one simulated draw, in which *i* subjects move through the specified stage structure over the interval $(s,t]$. The function takes the process' estimated stratified Cox model and uses the model's cumulative hazards, $\mathbf{A}(t)$, to simulate subjects' movement through the stages. As we mentioned in the main text, to estimate transition probabilities, the researcher must specify:

1. The subjects' current stage.
2. The time interval, *s* to *t.*
3. A covariate profile, listing the value at which every covariate should be held.

The function outputs one row for every unit increment between *s* and *t*, and the proportion of subjects occupying each stage at that point in time. For an example, Table 8 contains the output from one simulated draw for 110 subjects—the number of countries in Maeda's dataset—on the interval (0,40], in which all subjects start in Stage 3 and all subjects are more-developed countries with all other covariates at their median values. By the end of year 1, all 110 subjects remain in Stage 3; by the end of year 2, 7 subjects have transitioned into Stage 1 (6.3%) and the remaining 103 subjects are still in Stage 3 (93.6%); and so on.

TABLE 8. One `mssample` Draw

t	Stage 1	Stage 2	Stage 3	t	Stage 1	Stage 2	Stage 3
1	0.000	0.000	1.000	21	0.845	0.018	0.136
2	0.064	0.000	0.936	22	0.882	0.018	0.100
3	0.064	0.000	0.936	23	0.882	0.036	0.082
4	0.127	0.000	0.873	24	0.882	0.045	0.073
5	0.127	0.000	0.873	25	0.882	0.045	0.073
6	0.127	0.000	0.873	26	0.873	0.055	0.073
7	0.182	0.000	0.818	27	0.882	0.055	0.064
8	0.300	0.000	0.700	28	0.882	0.055	0.064
9	0.282	0.018	0.700	29	0.909	0.064	0.027
10	0.282	0.018	0.700	30	0.891	0.073	0.036
11	0.336	0.009	0.655	31	0.882	0.082	0.036
12	0.455	0.009	0.536	32	0.864	0.082	0.055
13	0.718	0.009	0.273	33	0.864	0.082	0.055
14	0.718	0.009	0.273	34	0.864	0.082	0.055
15	0.700	0.009	0.291	35	0.873	0.073	0.055
16	0.691	0.009	0.300	36	0.873	0.073	0.055
17	0.800	0.009	0.191	37	0.873	0.064	0.064
18	0.800	0.009	0.191	38	0.873	0.064	0.064
19	0.791	0.009	0.200	39	0.855	0.082	0.064
20	0.836	0.018	0.145	40	0.855	0.091	0.055

Second, we loop over `mssample` N number of times, storing the output from each draw. Our simulations from the main text use $N = 1000$. The loop gives us N observations for every unit increment between s and t—for (0,40], we get N observations for $t = 1$, N observations for $t = 2$, etc. Obtaining our final simulated transition probabilities simply becomes a matter of averaging. Specifically, the simulated transition probability for Stage g in time t will be equal to the proportion of subjects in Stage g at t averaged across all N simulations. We compute 95% confidence intervals for Stage g's transition probability in t by taking the appropriate percentiles (2.5, 97.5) across all of t's N simulations.

Appendix I:
Application – Simulated Data: Parameter Values

TABLE 9. Simulated Data's True Parameter Values

BASELINE HAZARDS (α_{gh_0})		β_x	
Parameter	*Value*	*Parameter*	*Value*
α_{12_0}	0.15	$\beta_{(1\to2)x}$	0.5
α_{13_0}	0.05	$\beta_{(1\to3)x}$	3.0
α_{21_0}	0.02	$\beta_{(2\to1)x}$	1.0
α_{23_0}	0.05	$\beta_{(2\to3)x}$	-1.0

NOTE: α_{gh_0} = baseline hazard of transitioning from Stage *g* to Stage *h*

Appendix J:

Non-Parametric vs. Semi-Parametric vs. Parametric Multi-state Models

In the main text, we introduce multi-state models, and specifically discuss how to estimate semi-parametric variants of these models. However, as with other models for event history data, there exist non-parametric, semi-parametric and parametric alternatives. In this appendix, we briefly review the differences between these alternatives, and discuss instances in which one alternative may be preferable to another.

I. Non-Parametric

Non-parametric multi-state models, as with any non-parametric survival model, entail estimating an event history model without covariates. Such models are non-parametric in that they do not parameterize the baseline hazard(s)—the baseline hazards are estimated using the data's observed failure times only.

We strongly encourage researchers to estimate their multi-state model non-parametrically to begin. Non-parametric estimation permits practitioners to direct their focus to the number of observed transitions between stages, ensuring that the dataset is formatted properly, and even doing some preliminary tests for baseline hazard equivalence.[59] The baseline hazard tests can be a useful first step in hypothesis testing, as they allow researchers to assess different transition rates between stages, which may be of interest. Further, more complex multi-state models with many covariate effects can have underlying specification issues that affect the estimation of sensible transition probabilities (Fiocco, Putter, and Houwelingen 2008). These specification issues appear in other econometric models as well. Examples include having little variation in an independent variable for a particular transition or having very few observed events for a transition. Running the non-parametric model first allows researchers to use it as a diagnostic.

[59] These tests should be rerun once the model has covariates.

II. Semi-Parametric

Semi-parametric multi-state models were the subject of the main text and therefore require little further elaboration. As with non-parametric models, semi-parametric models do not parameterize the baseline hazard(s), allowing them to vary based upon observed failures in the data. Semi-parametric models differ from non-parametric models in that they include some parameters in the form estimated covariate effects. The principle advantage of a semi-parametric modeling approach is its inherent flexibility—the researcher is not required to make any assumptions about the shape or rate of the baseline hazards.

III. Parametric

Parametric event history models have been widely employed in political science (e.g., Bennett and Stam 1996). Parametric event history models differ from non- and semi-parametric event history models in that they parameterize the baseline hazard(s) by making an assumption about the nature of time dependence in the data (i.e., the dependence's distribution). Common distributions in political science include the Weibull, exponential, log-normal, and others. Parametric event history models can be extended to a multi-state context, similar to how semi-parametric models are extended. In this instance, a separate baseline hazard may still be estimated for each transition in the data, but an assumption is made regarding the baseline hazards' distributional form. A number of software packages are available to estimate parametric multi-state models, including the `msm` package for `R` (Jackson 2011).

The biggest drawback of parametric event history models is that they may yield inefficient estimates at best if the distribution specified for the baseline hazard(s) is erroneous, and biased results at worse (Box-Steffensmeier and Jones 2004, 21–22). We mentioned this in the paper's introduction. Relatedly, political science researchers often do not have hypotheses about the form of baseline hazards' distribution; having such a hypothesis would be a reason to parameterize the baseline hazard.[60] In short, parameterizing the baseline hazard does not usually serve any *direct* inferential purpose, because our hypotheses are not about the baseline hazards' distribution, and the cost of making an incorrect parametric assumption is deleteriously high, as our estimate of *x*'s effect on the transition in question can be biased.

However, parametric survival models do have a few advantages. First, by parameterizing the baseline hazard, parametric models may yield more efficient estimates than semi-parametric alternatives. This efficiency gain is one of the central advantages of any parametric event history model. Second, parametric multi-state models, particularly the `msm` package, embrace panel-formatted data more fully than `mstate`; the latter prefers data in counting-process format, which can handle some aspects of panel data, but requires more data manipulation and cleaning to do so. Finally, by parameterizing the baseline hazard, it is easier to implement more advanced multi-state models. Hidden Markov models (HMMs) and misclassification models are two prominent examples. HMMs allow for unobserved stages, and model the transition between these latent stages; our multi-state model requires observed stages to be estimated. Misclassification models are an extension to HMMs, and are concerned with situations in which the data erroneously record a subject as occupying one stage, when the subject is truly in another. In the presence of misclassification, the transition probabilities can be biased, depending on the severity and type of misclassification (Jackson et al. 2003). Misclassification models correct for misclassification by modeling two processes: one for the classification process itself, and then the usual equation for the actual theoretical process of interest.

---

[60] Whether one should draw inferences from the baseline hazard has also been a matter of debate. For contrasting views, see Carter and Signorino (2010a, 2010b) and Beck (2010).

Appendix K:
Strategic Processes and Multi-state Models

Both multi-state models and empirical models of strategic interactions are similar in that both are interested in processes. Whether multi-state models can assess hypotheses regarding strategic behavior is a complex answer. By default, they cannot. Multi-state models assume each stage's transition events are independent of one another—the same assumption underlying all competing risks models.[61] By contrast, one of the hallmarks of strategic interactions (and empirically modeling them[62]) is the *non*-independence of event outcomes. However, empirical models for dependent competing risks do exist. Future work could possibly adapt these dependent CR strategies to a multi-state setting, which would put multi-state models in a better position to address hypotheses regarding strategic interactions.

[61] See main text's fn. 10.

[62] Examples of such models include strategic probit, strategic logit, and strategic backward induction (e.g., Bas, Signorino, and Walker 2008; Leemann 2014; Nieman 2015; Signorino 1999, 2003).

Appendix L:
Number of Transitions per Independent Variable

One potential limitation of multi-state event history models is that as the number of stages and transition types increase, the number of observed transitions of each type may diminish. As with any type of application, from logistic regression to simple Cox models, reducing the number of observed events raises the risk of biased coefficient estimates and may even prevent a model from converging. To our knowledge, there is no definitive rule of thumb for the minimum number of transitions of each type to ensure multi-state model results remain unbiased. However, Vittinghoff and McCulloh (2007) suggest that a minimum of 5-10 events per independent variable is usually sufficient for Cox models. Extending this recommendation to multi-state models implies that, for each unique strata included in the model, there should be at least 5-10 observed transitions of that type per transition-specific independent variable.

In situations in which an insufficient number of observed transitions exist within a given strata, it is possible to collapse two or more unique transitions together, thereby increasing the number of observed transitions within that particular strata. For instance, if there are 9 observed transitions from 1→3 and 8 transitions from 2→3, it is possible to stratify the baseline hazards for each transition, but only a single transition-specific independent variable would be advisable for each transition. Pooling the transitions together would yield 17 events in the new strata, allowing two additional covariates to be included instead of one. A related strategy in the presence of relatively rare transitions entails also holding these transitions' covariate effects constant (i.e., no transition-specific covariates). This strategy can also help with convergence issues and improve the estimates' precision. Box-Steffensmeier and Zorn (2002, 1079) offer similar advice in the context of conditional models for repeated events, where the number of subjects at risk of experiencing the event may become increasingly small as the number of previous events increases.[63] As discussed in the main text, decisions about collapsing covariate effects should not be made lightly, and should derive from a combination of theory and statistical fit.

[63] For instance, many House incumbents may get reelected once or twice, but fewer get reelected for a 15^{th} or 16^{th} time.

Appendices Works Cited

Allison, Paul D. 2014. *Event History and Survival Analysis*. 2nd ed. Thousand Oaks, CA: Sage.

Andersen, Per Kragh, and Maja Pohar Perme. 2013. "Multistate Models." In *Handbook of Survival Analysis*, eds. John P. Klein, Hans C. van Houwelingen, Joseph G. Ibrahim, and Thomas H. Scheike. Boca Raton, FL: Chapman and Hall/CRC, pp. 417–439.

Bas, Muhammet Ali, Curtis S. Signorino, and Robert W. Walker. 2008. "Statistical Backwards Induction: A Simple Method for Estimating Recursive Strategic Models." *Political Analysis* 16 (1): 21–40.

Beck, Nathaniel. 2010. "Time is Not A Theoretical Variable." *Political Analysis* 18 (3): 293–294.

Bennett, D. Scott, and Allan C. Stam. 1996. "The Duration of Interstate Wars, 1816-1985." *American Political Science Review* 90 (2): 239–257.

Beyersmann, Jan, Arthur Allignol, and Martin Schumacher. 2011. *Competing Risks and Multistate Models with R*. New York: Springer.

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

Box-Steffensmeier, Janet M., and Christopher J.W. Zorn. 2002. "Duration Models for Repeated Events." *Journal of Politics* 64 (4): 1069–1094.

Carter, David B., and Curtis S. Signorino. 2010a. "Back to the Future: Modeling Time Dependence in Binary Data." *Political Analysis* 18 (3): 271–292.

------. 2010b. "Reply to 'Time is Not A Theoretical Variable.'" *Political Analysis* 18 (3): 295–296.

Enders, Walter. 2009. *Applied Econometric Time Series*. 3rd ed. New York: Wiley.

Fiocco, Marta, Hein Putter, and Hans C. van Houwelingen. 2008. "Reduced-Rank Proportional Hazards Regression and Simulation-Based Prediction for Multi-State Models." *Statistics in Medicine* 27 (21): 4340–4358.

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

Jackson, Christopher H. 2011. "Multi-State Models for Panel Data: The msm Package for R." *Journal of Statistical Software* 38 (8): 1–28.

Jackson, Christopher H., Linda D. Sharples, Simon G. Thompson, Stephen W. Duffy, and Elisabeth Couto. 2003. "Multistate Markov Models for Disease Progression with Classification Error." *The Statistician* 52 (2): 193–209.

Keele, Luke, and Nathan J. Kelly. 2006. "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables." *Political Analysis* 14 (2): 186–205.

Lebo, Matthew J., and Will H. Moore. 2003. "Dynamic Foreign Policy Behavior." *Journal of Conflict Resolution* 47 (1): 13–32.

Leemann, Lucas. 2014. "Strategy and Sample Selection: A Strategic Selection Estimator." *Political Analysis* 22 (3): 374–397.

Maddala, G. S., and In-Moo Kim. 1998. *Unit Roots, Cointegration, and Structural Change*. Cambridge: Cambridge University Press.

McGrath, Liam F. 2015. "Estimating Onsets of Binary Events in Panel Data." *Political Analysis* 23 (4): 534–549.

Meira-Machado, Luís, Jacobo de Uña-Álvarez, and Carmen Cadarso-Suárez. 2006. "Nonparametric Estimation of Transition Probabilities in a Non-Markov Illness–Death Model." *Lifetime Data Analysis* 12 (3): 325–344.

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

Nieman, Mark David. 2015. "Statistical Analysis of Strategic Interaction with Unobserved Player Actions: Introducing a Strategic Probit with Partial Observability." *Political Analysis* 23 (3): 429–448.

Piger, Jeremy. 2011. "Econometrics: Models of Regime Changes." In *Complex Systems in Finance and Econometrics*, ed. Robert A. Meyers. New York: Springer, pp. 190–202.

Potter, Simon. 1999. "Nonlinear Time Series Modelling: An Introduction." *Journal of Economic Surveys* 13 (5): 505–528.

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

Signorino, Curtis S. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *The American Political Science Review* 93 (2): 279–297.

------. 2003. "Structure and Uncertainty in Discrete Choice Models." *Political Analysis* 11 (4): 316–344.

Spirling, Arthur. 2007. "Bayesian Approaches for Limited Dependent Variable Change Point Problems." *Political Analysis* 15 (4): 387–405.

Vittinghoff, Eric, and Charles E. McCulloch. 2007. "Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression." *American Journal of Epidemiology* 165 (6): 710–718.

Wawro, Gregory. 2002. "Estimating Dynamic Panel Data Models in Political Science." *Political Analysis* 10 (1): 25–48.

Wreede, Liesbeth C. de, Marta Fiocco, and Hein Putter. 2010. "The mstate Package for Estimation and Prediction in Non- and Semi-Parametric Multi-State and Competing Risks Models." *Computer Methods and Programs in Biomedicine* 99 (3): 261–274.