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

May 7, 2015

Prepared for the Visions in Methodology conference
Lexington, KY
May 13-15, 2015

Preliminary draft.
Please do not circulate further without permission

Abstract:
Many political processes consist of a series of theoretically meaningful transitions across discrete phases. While regime-switching models allow us to empirically assess hypotheses about transitions between phases in some contexts, there have been relatively few attempts to extend such models to the study of durations. Yet, political scientists are often theoretically interested in studying not just transitions between phases, but also the duration that subjects spend within phases. We introduce the multi-state survival model to political scientists, which is capable of modeling precisely this type of situation. The model is appealing because of its ability to model multiple forms of causal complexity that unfold over time. In particular, we highlight three attractive features of the multi-state model: its stratification of baseline hazards, its transition-specific covariate effects, and its ability to estimate overall transition probabilities. We provide two illustrative examples from different subfields to illustrate the model's features.

---

* The authors' names appear in reverse alphabetical order. We bear sole responsibility for any remaining errors and shortcomings. All analyses are performed using R 3.1.2.

The notion of "change over time" is a prominent part of many political science research agendas. What influences democratization? What explains shifts in the American electorate? How do financial crises spread to multiple countries? In all these examples, the passage of time provides an opportunity for some outcome to exhibit variation, particularly within a specific case. Researchers then exploit this variation to help evaluate their theories, often using longitudinal and time-series methods.

Many of these processes evolve through a series of interdependent phases/stages, begging a question about how long a subject remains in a stage before transitioning to another. In parliamentary systems, governments must first form, then govern (Chiba, Martin, and Stevenson 2015; Hays and Kachi 2009; King et al. 1990). How long does it take to form the government? Once the government is formed, how long does it survive before being dissolved? In international relations, a diplomatic dispute between states first begins, potentially escalates, becomes militarized, and then enters a post-militarization peace (Diehl 2006, 200). How long before a dispute militarizes? And, once it militarizes, how long before the militarization ends? How long before the dispute is resolved at all?

Despite our typical interest in the *specific* duration of each stage, we also have an implicit interest in these processes *as a whole*. The stages' interconnectivity implies substantively relevant byplay that could enrich our understanding of these processes. For instance, the presence of different stages implies that the same covariate could have different effects, depending on the stage in question. If we are interested in assessing this covariate's effect on 'time in stage', ignoring the different stages could produce biased estimates.

Yet, to date, our standard empirical tests are ill-equipped to juggle all of these balls at once.[1] Regime switching models, as a class of models, are superb at modeling stage-specific covariate effects, and can handle very complicated stage sequences, but usually do not focus on durations. Standard survival models can speak to durations, but are less adroit at handling many stage-specific covariate effects. They also cannot handle complicated stage sequences like recursiveness, where a subject can

[1] For some intriguing partial exceptions, see Chiba, Metternich, and Ward (forthcoming).

occupy the same stage more than once. A similar truth holds for logit/probit and their respective multinomial variants. These models have a limited ability to capture more complex stage sequences, and handling many stage-specific covariate effects can be cumbersome. However, the models can accommodate durations by adding time counters as regressors (Beck, Katz, and Tucker 1998).

As a consequence of our standard tests' limitations, researchers tend to focus on only one transition within a process (e.g., militarization in a dispute). However, in doing so, researchers lose the ability to say anything holistic about the process, particularly when it comes to the probability of transitioning into a particular stage at a point in time. What, for instance, is the probability of a dispute being resolved at some *t*, given that states can repeatedly militarize a dispute, and that states can repeatedly try to resolve the dispute through peaceful negotiations?

How, then, should we investigate claims about durations in stage-based processes? We make the novel suggestion that survival models *are* capable of investigating claims about stage-based processes. We introduce a more advanced model, the multi-state survival model, to political science (Therneau and Grambsch 2000). Multi-state models are stratified Cox models in which covariates can have different effects, depending on the transition in question.[2] The model is therefore capable of handling durations, complex stage sequences, and many stage-specific covariate effects. Multi-state models estimate all of this in a unified framework, which permits practitioners to compute overall transition probabilities using information about every transition within the process, instead of only one transition within it. We think this should be particularly attractive to political scientists, because of its rich theoretical potential.

Our discussion proceeds in four parts. We begin by introducing the model itself. Second, we highlight the model's attractive features. Third, we provide some illustrative applications, to show how the model works, and the types of inferences we can draw from it. The fourth and final section concludes.

---

[2] Transitions describe instances in which a subject moves *from* one stage *into* another one. Notice how they consist of 'from-to' stage pairings.

I. What are multi-state models?

A multi-state survival model, synonymously referred to as a multi-state event history model, is an econometric estimator capable of modeling a duration process that is comprised of multiple 'stages' (Therneau and Grambsch 2000). Our interest is often in understanding (a) when transitions between stages will occur, (b) the probability of the transitions, and (c) what covariates increase or decrease these transition probabilities. Multi-state models have been used to explore causes of death among Norwegian citizens (Vollset, Tverdal, and Gjessing 2006), bone marrow recipients' health (Putter, Fiocco, and Geskus 2007, 2417–2422), and individuals' cohabitation patterns (Mills 2011). However, their use in political science has been very rare.[3]

Despite their infrequent use in political science, multi-state models are built from methodological pieces that *are* familiar to political scientists. Accordingly, we introduce multi-state models from the ground up, using these pieces—we begin with simple survival models, move to competing risks models, and then finally arrive at multi-state models.

A. Basic Survival Models

Say that we are interested in the occurrence of a particular event, with a specific interest in how long it takes subject *i* to experience this event. Survival models, also known as duration models or event history models, are well-suited to answering questions of this form. They are interested in modeling the event's hazard rate, which (loosely) expresses the probability of *i* experiencing the event in *t*, contingent upon *i* still being at risk for experiencing the event in *t* (Allison 1984).[4] The hazard itself is unobserved, but we suppose that it is a function of *i*'s "time at risk" for the event, permitting us to model the hazard using the observed duration. This duration is typically defined as how much time passes between the first

[3] Exceptions include Jones (2013), Jones and Metzger (2015), and Jones and Mattiacci (2015).

[4] In truth, hazards are not unconditional probabilities, per se. They represent the instantaneous risk of failure. For continuous-time durations, they are conditional probabilities (Aalen, Borgan, and Gjessing 2008, 5–6). A hazard rate can be larger than one, for instance (Cleves et al. 2010, 7–8). Describing hazards as being probability-like simply helps, for expositional purposes.

period in which *i* could have experienced the event, and the period in which *i* did experience the event.[5] If we begin "counting" from 0 in the first period, then *t* represents the amount of time that a subject has been at risk for the event.

A number of well-known, basic survival models are available to us. On the one hand, parametric survival models assume a specific functional form for the baseline hazard rate (Box-Steffensmeier and Jones 2004, chap. 3), where the baseline hazard rate expresses the event's hazard rate when the covariates are equal to zero. Examples include the exponential, Weibull, gamma, and log-normal models.

On the other hand, semi-parametric survival models do not make any parametric assumptions about the baseline hazard rate. Instead, they parameterize only the covariates' relationship with the hazard, and estimate these coefficient values using partial-likelihood methods (Box-Steffensmeier and Jones 2004, chap. 4). The Cox proportional hazards model is the quintessential semi-parametric survival model, and is the building block for our more advanced multi-state model. The Cox's hazard rate is expressed as (Box-Steffensmeier and Jones 2004, 48):

$$\alpha(t) = \alpha_0(t)e^{\beta^T Z} \qquad 1$$

where $\alpha(t)$ is the hazard of the event occurring in time t, $\alpha_0(t)$ represents the baseline hazard rate in t, Z is a vector of covariates, β is the vector of coefficients, and T is the transpose of the vector.[6]

[Insert Figure 1 about here]

B. Competing Risks Models

Basic survival models assume that subjects are only at risk of experiencing one event. What happens if subjects are at risk of experiencing *multiple* events? Competing risks (CR) models can handle this additional wrinkle. CR models are a special type of multi-state model, which makes the former useful for beginning to explain the latter's features.

---

[5] For different ways to define the start and end points for a duration of interest, see Skrondal and Rabe-Hesketh (2004, 373–376).

[6] Notation adopted from Wreede, Fiocco, and Putter (2010).

Formally, CR models extend standard Cox survival models. They are still interested in subject *i* being at risk for experiencing an event, and model how long the event takes to occur. The key difference between standard Cox models and CR models is that subject *i* is at risk of experiencing *two or more such events*. These multiple events are referred to as "transitions," in the parlance of multi-state models. The implication is that there are multiple ways in which *i*'s time at risk can end. Figure 1(b) visually depicts the CR scenario, while Figure 1(a) depicts the standard Cox scenario. In both panels, Stage 1 represents a subject's initial 'time at risk'. In 1(a), there is only one way for a subject to exit Stage 1—a transition into Stage 2. By contrast, in 1(b), there are two possible ways for a subject to exit Stage 1—a transition into Stage 2, or a transition into Stage 3.

Recognizing that there are multiple transitions out of a specific stage is important. If we pool all the stage's exiting transitions together, we are implicitly assuming that each transition's data-generating process (DGP) is identical. A covariate would therefore have the same effect on every transition. If the transitions have different DGPs, though, a pooled-transition model would produce biased estimates. The estimates would equal the average effect of the covariate, across all the transitions.

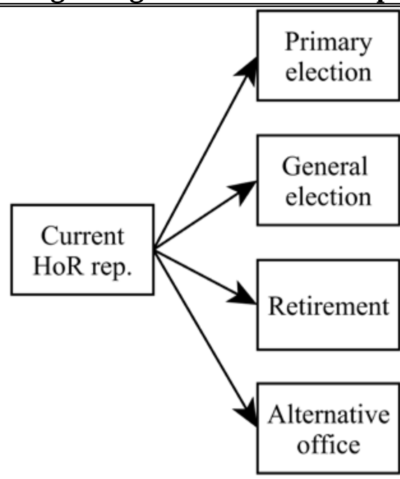
In a classic CR setup, all observations (1) begin in the same stage, (2) are simultaneously at risk of experiencing two or more transitions, and (3) after experiencing one of the transitions, an observation is no longer at risk of experiencing *any* transitions (Box-Steffensmeier and Jones 2004).[7] A classic application pertains to the legislative careers of US House representatives (2004, 169–172). We depict this process' stage diagram in Figure 2. Any incumbent representative's tenure in the House will end, eventually. However, there are several ways the incumbent could leave office. The representative could:

1. Be defeated in a primary election
2. Be defeated in the general election
3. Choose to retire
4. Seek alternative office (e.g., Senate, gubernatorial, cabinet appointment)

[7] A classic CR setup also assumes that the different events are independent of one another. Multi-state models make the same assumption. For models that explore dependent competing risks, see Gordon (2002) and Fukumoto (2009).

Once a representative has experienced one of these four events, s/he has "exited the risk set", and is no longer at risk of experiencing the other three. A representative who retires, for instance, would no longer be at risk of exiting the House via electoral defeat.

FIGURE 2. Stage Diagram – House of Representatives



A classic CR model recognizes the different possible transitions out of a risk set, and estimates an equation for each transition. For semi-parametric survival models, Cox models and Fine-Gray subhazard models (Fine and Gray 1999) are the most common estimators.[8] CR's major modeling strength is that it permits a covariate's effect to vary across transitions. Doing so guards against the biased estimates that would potentially result from pooling all the transitions. For example, a covariate that *appreciably* increases the probability of primary election defeat may only *slightly* increase the probability of general election defeat. A CR model would detect this difference, whereas standard Cox model with pooled transitions would not.

Yet, a classic CR model is limited in its ability to model more complex situations. It is primarily focused on transitions out of the starting stage. The model does not speak to what happens to subjects after they transition out of Stage 1 and into, e.g., Stage 2 or Stage 3 (Figure 1(b)). We can imagine situations in which this information would be substantively useful. An additional implication is that CR models cannot handle situations in which, after a subject experiences one event, the subject is still at risk

[8] All the parametric models from the previous section can also handle competing risks. Semi-parametric approaches are simply more common, because of their more flexible assumption regarding the baseline hazard.

of experiencing other events. An ongoing territorial dispute that militarizes, for instance, may still experience peaceful negotiations (Jones and Metzger 2015).

C. Multi-state Models

Multi-state models take a holistic approach to a process. They "extend the analysis to what happens after the first [transition] event," allowing researchers to model how a subject moves through several stages (Putter, Fiocco, and Geskus 2007, 2390). The implication is that multi-state models permit *multiple* risk sets, vs. CR's single risk set. Consequently, they are sufficiently flexible to model any number of possible process structures, using a single framework. They can capture situations in which events occur sequentially, repeatedly, or any combination thereof (Putter, Fiocco, and Geskus 2007; Therneau and Grambsch 2000). In short, we can use multi-state models to estimate a process with *any* of the stage structures depicted in Figure 1, whereas classic CR models can only handle the first two panels, and a standard Cox model could only handle the first.

The premise of multi-state models is simple: a subject transitioning *out* of one stage must be transitioning *into* another one. Rather than dropping the subject after this first transition (like classic CR does), multi-state models consider what new transitions the subject is now at risk of experiencing. This is how multi-state models are comprised of multiple risk sets. More specifically, multi-state models use stages to define different risk sets, since subject *i*'s current stage determines which transitions *i* is at risk of experiencing. For a concrete example, take a complex process like Figure 1(f). A subject in Stage 1 is at risk of experiencing two transitions: one into Stage 2, and one into Stage 3. By contrast, a subject in Stage 2 is at risk of experiencing one transition, into Stage 4.

Multi-state models can be estimated as stratified Cox models, which differ from a standard Cox model in two key respects.[9] First, the underlying *baseline hazard is stratified* for each of the possible transitions within the process. A separate baseline hazard, $\alpha_{q0}(t)$, is estimated for each possible

[9] Classic CR models are different from standard Cox models in the same two ways, since classic CR models are just a specific example of a (simple) multi-state model.

transition q, where t continues to refer to the amount of time that a subject has been at risk.[10] By contrast, standard Cox models only estimate one baseline hazard, $\alpha_0(t)$; there are no q subscripts.

Second, multi-state models include *transition-specific covariates*, Z_q, where q, as above, indexes every possible transition in a process. Doing so allows each variable to have a different effect, depending on the transition in question. For example, some x might decrease the risk of transitioning from Stage 1 to Stage 2, but might increase the risk of transitioning from Stage 2 to Stage 3. Allowing for such differences is important, as it allows for transition-specific covariate effects. By contrast, a typical standard Cox model can only accommodate one transition, making the transition-specific designation irrelevant.

Thus, the hazard rate for a multi-state model is given by:[11]

$$\alpha_q(t) = \alpha_{q_0}(t)e^{\beta^T Z_q} \qquad 2$$

Given the hazard rate identified above, cumulative transition hazards may be estimated as:

$$A_q(t) = \int_0^t \alpha_q(u)du \qquad 3$$

and aggregated into an S x S matrix, $\mathbf{A}(t)$, where S is the number of possible stages within the multi-state model, and u denotes all event times within some time interval $(s,t]$.[12] For the US House legislator example, S would be equal to 5—(1) in office, (2) primary election defeat, (3) general election defeat, (4) retirement, and (5) assuming an alternative office.

Cumulative transition hazards are relevant, because they permit us to calculate transition probabilities. Specifically, we can estimate a transition probability matrix, $P(s,t)$, as:

$$P(s,t) = \prod_{u \in (s,t]} (\mathbf{I} + \Delta\mathbf{A}(u)) \qquad 4$$

---

[10] However, these baseline hazards need not be modeled separately, should theory or statistical tests indicate that two or more of them are equal. We discuss this further in a later section.

[11] Contrast this with the hazard rate for a standard Cox model; see Equation 1.

[12] The stages are numbered purely for organizational purposes.

where $(s,t]$ denotes the time interval. The individual elements of the matrix $P(s,t)$ are the probability of transitioning from each stage to every other stage within the time interval $(s,t]$.[13] For the US House example, element $P_{1,2}(s,t)$ would denote the probability of a legislator transitioning from Stage 1 (in office) to Stage 2 (defeat in the primaries), within the time interval s to t. Importantly, these transition probabilities will vary over time, because the hazards on which they are based vary as well. This means that, holding all else constant, the probability of a particular transition occurring may be substantially different at time t than it is at time $t + 5$.

To begin concluding, the above discussion 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:

$$y_t = \beta_{0_{S_t}} + \beta_{1_{S_t}} x_1 + \beta_{2_{S_t}} x_2 + \cdots + \beta_{k_{S_t}} x_k \qquad 5$$

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.[14] Equation 5 is the same general form taken by 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.[15] Multi-state models pertain to durations, a quantity that is less talked about in the regime-switching context. Additionally, in our multi-state model, S is known and observed by the researcher. Jackson (2011) discusses possible

---

[13] Transitions that are impossible, either realistically or theoretically, are held at 0.

[14] When conceptualizing regime switching models, it is common to think of the β's as changing across regimes. However, these are not the only parameters that could change. Other examples include autoregressive parameters, and the covariance of y_t's idiosyncratic error (Weskamp and Höchstötter 2010, 17).

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

multi-state extensions for applications in which a process' stages are unobserved, using parametric survival models. 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.

II. Why are multi-state models useful?

Multi-state models are an extension of the familiar Cox survival model, and as such, share many of the beneficial properties of Cox models and other extensions familiar to political scientists, such as competing risks models. From this common foundation, multi-state models present a highly flexible approach to the study of political processes that unfold over time across a series of possible transitions. Specifically, the use of a multi-state modeling framework introduces three innovations to the use of survival models in political science, several of which build upon current practices that, though used often, have not been applied more generally. We consider each of these innovations in turn.

A. Stratification of Baseline Hazards

One of the primary advantages in the use of multi-state models is the flexibility that they afford researchers to model any number and sequence of events that are deemed to be theoretically or substantively meaningful. In order to accommodate these varied event sequences, a multi-state modeling strategy allows the researcher to stratify the baseline hazard for each of the different transitions in the model. In practice, this simply means that the underlying rate at which one type of event occurs is allowed to vary from the underlying rate at which an event of a different type occurs. This type of stratification is familiar to researchers that employ competing risk models (Box-Steffensmeier and Jones 2004), depicted visually in Figure 1(b). In modeling this situation, the baseline hazard is stratified by each possible transition to reflect the possibility that the underlying rate at which transitions from Stage 1 to Stage 2 occur may vary from the rate at which transitions from Stage 1 to Stage 3 occur.

Stratification is also a prevalent strategy when dealing with repeated events, depicted visually in Figure 1(c). The underlying rate at which subjects experience the first event may differ from the rate at which subjects experience a second or third event, if, for example, experiencing a first event makes subjects more likely to experience subsequent events (Box-Steffensmeier and Zorn 2002). Stratification in the context of repeated events also underscores a related issue, which is that not all subjects are necessarily at risk for all transitions simultaneously. Rather, transitions may only occur sequentially, such that some subjects only become at risk for a particular transition *after* experiencing a previous event. In the context of repeated events of the same type, this is straightforward. A subject is only at risk of experiencing a second event after it has experienced a first event (as in conditional models of repeated events; see Prentice, Williams, and Peterson 1981). This same principle can generalize to situations in which there may not only be repeated events of the same type, but also different events that occur in a sequence. Figure 1(c) illustrates this more general situation, where all subjects begin in Stage 1 and are at risk of a transition to Stage 2. However, subjects *only* become at risk of a transition to Stage 3 once they have already transitioned into Stage 2.

By employing precisely this stratification approach, multi-state models are capable of modeling many more complex sequences of events that may combine one or more of the characteristics of competing risks and repeated events models discussed above. As Figures 1(d-f) reflect, stratification of the baseline hazard allows for the researcher to differentiate between many different types of event sequences that may arise in their data. The determination of the "appropriate" number of transitions and their sequence is largely a matter of theoretical and substantive concern, depending on the particular situation to which multi-state models are being applied. Nevertheless, it is also straightforward to use conventional goodness-of-fit tests to determine whether two or more baseline hazards are statistically different from one another, or whether they should be collapsed to a single transition. For example, in Figure 1(c), it is possible that the timing of the first event is significantly different from the timing of the second event, but that the timing of all subsequent events is not statistically different.

B. Unique Coefficient Estimates across Transitions

A second, but related, advantage afforded by multi-state models is the ability to estimate unique coefficient effects across each of the specified transitions in the model, allowing a researcher to determine whether the same covariate of interest exerts a different effect at different stages of a larger process. Again, this advantage is similar, in a limited sense, to a classic competing risks model in which the determinants of one stage of interest are allowed to vary from the determinants of another stage of interest. For example, consider the initial transitions in Figure 1(f), in which subjects located in Stage 1 are simultaneously at risk of two transitions (events): a transition into Stage 2, and a transition into Stage 3. As in a competing risks model, the baseline hazard of each transition is allowed to vary, but so too are the effects of the independent variables, such that the same covariate of interest may exert a different effect on the timing of one transition from another.

However, the use of *transition-specific covariates* is not limited to only a competing risks situation. It can be extended to each of the specified transitions in the model. For example, it is possible to examine whether the occurrence of intermediate events in a process alter the determinants of the same stage of interest. Consider the possible transitions into Stage 2 depicted in Figure 1(f). In contrast to the familiar competing risks model, there are two separate transitions that are possible into a single stage of interest, depending on whether a subject is directly transitioning into Stage 2 from the initial stage, Stage 1, or whether the subject has experienced an intermediate transition into Stage 3.[17] In this context, multi-state models allow for the estimation of unique covariates of interest for each of these two transitions, as it may be the case that by experiencing an intermediate event in the form of Stage 3, the determinants of transitioning into Stage 2 have fundamentally changed. That is, the effect of a covariate, x, on the timing of a transition into Stage 2 may be equal to $\beta_{(1\rightarrow2)x}$ if a subject is currently in Stage 1, and $\beta_{(3\rightarrow2)x}$ if a

[17] In many ways this is similar to a probit or logit, which would allow for the estimation of distinct coefficients depending on the occurrence of an intermediate event via interactions (Brambor, Clark, and Golder 2006). However, the advantage of the multi-state model in this context is that it allows a much greater degree of flexibility in terms of the number of transitions estimated, the order in which they are experienced, and the number of subsequent events for which a subject is at risk.

subject is instead currently located in Stage 3. The use of transition-specific covariates to estimate these unique covariate effects is possible across *each* of the 5 transitions depicted in Figure 1(f).

As with the stratification of baseline hazards for different transitions, multi-state models are exceptionally flexible in the specification of unique covariate effects. As such, the decision regarding how many unique covariate effects ought to be estimated in any given context is largely a matter of theory. It is entirely possible to estimate a unique coefficient for each covariate in the model across each of the transitions, provided that there are an adequate number of observed transitions of each type. However, in many contexts, it may be inappropriate to estimate a unique coefficient across each of the transitions in a model, as a covariate may exert a similar effect across one or more transitions. For example, if researchers suspect that some x exerts the same effect on two or more transitions in a multi-state model, then a single coefficient for x may be estimated for those transitions, thus holding the effect of x constant across each of those transitions. Wald tests for the equivalence of one or more coefficient estimates can aid in determining whether and how many unique coefficient estimates are appropriate in a particular application (Greene 2012, 113–121). This can be done either by conducting pairwise comparisons of coefficient estimates across transitions—for example, testing whether $\beta_{(1\rightarrow 2)x}$ is significantly different from $\beta_{(3\rightarrow 2)x}$—or by conducting joint tests of significance for whether $\boldsymbol{\beta}_{\mathbf{(1\rightarrow 2)}}$ are significantly different from $\boldsymbol{\beta}_{\mathbf{(3\rightarrow 2)}}$ (Therneau and Grambsch 2000, 226).

C. Transition Probabilities

While stratification of baseline hazards and unique covariate estimates across transitions allow for more precise modeling of the distinct transitions that constitute a larger process, they nevertheless are ill-suited, on their own, to making more systematic inferences about a political process as a whole. For example, if we consider Stage 4 in Figure 1(f) as the final outcome of interest, focusing solely on the transition from Stage 2 to Stage 4 would limit our understanding of the political process through which that final transition arises. Multiple event sequences could lead to the transition from Stage 2 to Stage 4. This is especially true if, for theoretical or substantive reasons, we are interested in how subjects move from the initial stage, Stage 1, to the final stage of the process, Stage 4. Focusing on the risk of each individual transition in isolation can provide, at best, a piecemeal understanding of how the political process unfolds. In order to understand this process in its entirety, and make inferences about it as a whole, it is necessary to aggregate the risk of each individual transition in the process.

The estimation of transition probabilities from multi-state models address this concern by providing an estimate of the probability of a subject occupying each stage in the model at time t, $t + 1$, $t + 2$ and so on (Wreede, Fiocco, and Putter 2010, 2011). Transition probabilities help to overcome the concern noted above about the possibility of multiple paths through which a subject could arrive at a particular stage of interest, as they take into account the probability of both direct and indirect transitions through the use of a product integral. In other words, if we are interested in the probability of a subject transitioning from Stage 3 to Stage 4 over some period of time, the transition probability estimate would take into account *each* of the possible paths through which a subject could arrive at Stage 4, given that it occupies Stage 3 in the present. For example, a subject could transition from Stage 3 to Stage 2 and then to Stage 4, or it could move from Stage 3 to Stage 1 and then to Stage 4, along with a number of other possible paths given the recursive nature of Figure 1(f).

As this example suggests, transition probability estimates are based on three key pieces of information:

1. The *stage a subject currently occupies*. Given that the baseline hazard of each transition in the model is allowed to vary, the stage that a subject currently occupies may have a substantial impact on the probability that it arrives at a subsequent stage of interest. It may be the case that, for example, transitions from Stage 1 to Stage 2 in Figure 1(f) occur relatively quickly, whereas transitions from Stage 3 to Stage 2 are quite protracted. If this were the case, the probability of a subject occupying Stage 2 would differ dramatically depending on the subject's stage in the present. Similarly, the probability of a subject occupying Stage 4 may also vary dramatically depending on the subject's

current stage, as it is necessary to arrive in Stage 2 before a subject is at risk of transitioning to Stage 4.

2. *The time frame for which* the transition probabilities are to be estimated. For instance, are the transition probability estimates to begin at the initial time under study, or only after some time has elapsed?

3. *A covariate profile of interest* by fixing each of the covariates in the model at a particular value, similar to estimating predicted quantities of interest from other estimators. In so doing, it becomes possible to evaluate the effect of a covariate of interest not only on a particular transition, but on the process as a whole. This becomes especially important if a covariate is found to exert opposite effects on different transitions (e.g., exerting a positive effect on one transition, and a negative effect on another). By estimating transition probabilities, it is possible to evaluate the *net* effect of a particular covariate on the process as a whole.

III. Applications

We reexamine several datasets, to show the utility of multi-state models. Model estimation is straightforward, as multi-state models are extensions of semi-parametric Cox models. As such, they may be readily estimated using widely used statistical software packages such as Stata and `R` once the data are structured properly.[18] The `mstate` package in `R` (Wreede, Fiocco, and Putter 2010, 2011) is specifically designed to facilitate the estimation and interpretation of multi-state models. It also has a number of utilities that aid with data manipulation, and most importantly, provide the ability to directly estimate

[18] For dataset organization details, see Jones and Metzger (2015, Appendix A), and Wreede, Fiocco, and Putter (2010). *VIM note, to discussant: we've attached the aforementioned appendix to the end of this paper's PDF, in case you have an interest.*

transition probabilities.[19] We rely on this package to estimate transition probabilities in each of the applications below.

A. Maeda: Modes of Democratic Breakdown

Maeda (2010) examines democratic reversals worldwide after 1950. The central theoretical contribution of the paper is that there are multiple ways in which a democratic regime may end. A democratic regime could be "exogenously" terminated, from outside the government itself (e.g., military coups). It could also be "endogenously" terminated, from inside the government (e.g., self-coups).[20] Maeda shows that his covariates of interest have different effects, depending on the "mode of democratic breakdown" (2010, 1129).

Maeda is specifically interested in democratic reversals. However, if we are ultimately interested in what factors contribute to or hinder the presence of democratic regimes in states, it also makes sense to look at the entire democracy breakdown-restoration process, instead of one piece of it. Thus, we reexamine Maeda's dataset, but in addition to democratic breakdown, we simultaneously consider "democratic restorations" (see Figure 3). We ask: If a democracy becomes a non-democracy, does it revert back? If so, how long before it does? Table 1 contains information about all 82 transitions in Maeda's expanded dataset, based on what stage the state is transitioning *from* (the current stage) and what stage it is transitioning *to* (the next stage).

[19] `mstate` is capable of simulating transition probabilities as well.

[20] Maeda also has 6 cases of "other" democratic terminations, where the source was neither inside the government, nor outside of it. E.g.) France's shift from the Fourth to Fifth Republic. For more details, see Maeda (2010, 1135, fn. 13).

FIGURE 3. Stage Diagram - Maeda

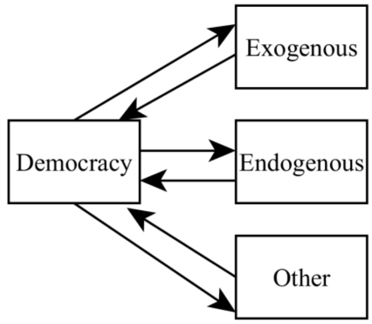


TABLE 1. Observed Transition Frequencies – Maeda

Current Stage	*Next Stage* Democracy	Exog.	Endog.	Other	TOTAL (for %s)
Democracy	–	24	18	6	48
(Row Total %)		(17.4%)	(13.0%)	(4.3%)	(34.7%)
Exogenous	15	–	–	–	15
(Row Total %)	(62.5%)				(62.5%)
Endogenous	14	–	–	–	14
(Row Total %)	(66.7%)				(66.7%)
Other	5	–	–	–	5
(Row Total %)	(83.3%)				(83.3%)
TOTAL	34	24	18	6	82

"TOTAL" column contains the total number of transitions across the row, representing the number of transitions from the current stage to another stage. The percentage calculations use the number of overall states in the current stage as the denominator. They do not sum to 100% because of right censoring. Cells containing a dash indicate impossible transitions in our model.

1. NON-PARAMETRIC

We begin our analysis by considering a non-parametric multi-state model, which estimates a distinct baseline hazard for each of the transitions in Figure 2.[21] This model, simply, captures the underlying rate at which each of the respective transitions occurs in the data without including any covariates. Though researchers are typically interested in the effect of covariates of interest on the timing

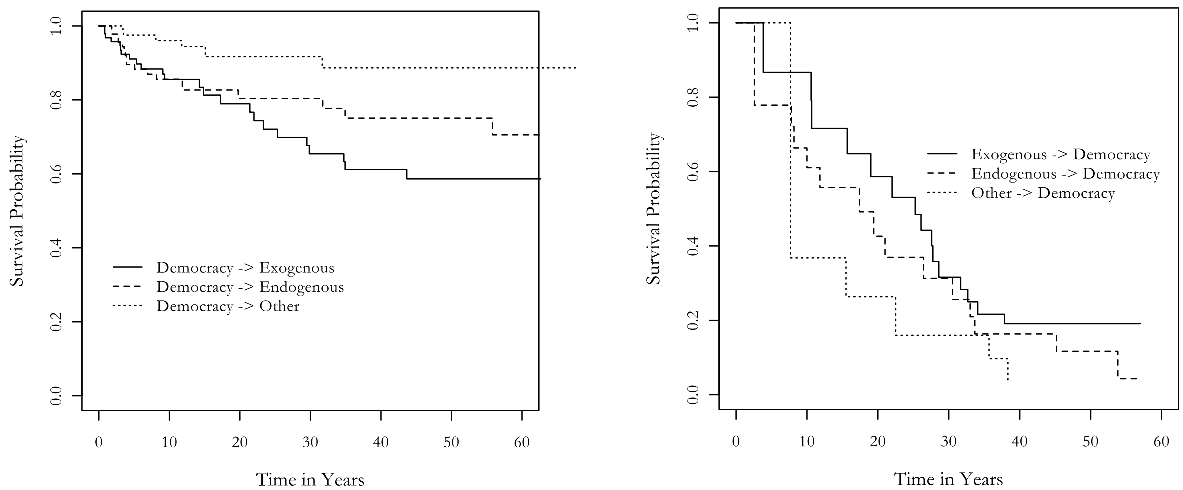
[21] We could have also estimated semi-parametric models for Maeda's dataset, as we do with our next application. For arbitrary explication purposes, we choose to start with a non-parametric model for Maeda (and not to report the semi-parametric results, for now), and to start with a semi-parametric model for the next application (and not to report the non-parametric results there). We could have just as easily swapped the applications.

of these transitions, a non-parametric model is a useful way to begin because it straightforwardly presents the ways in which multi-state models are similar to, *and distinct from*, other survival models.

Figure 4 plots the non-parametric survival estimates from this model. Figure 4(a), replicating Maeda, depicts the likelihood that a democratic state will breakdown either due to exogenous, endogenous, or some other reason. The figure shows that the likelihood that democratic regimes will fail due to exogenous and endogenous terminations is largely the same for the first 20 years of a democratic regime's existence. However, after 20 years, democratic regimes become significantly more likely to experience exogenous terminations (solid line) than endogenous terminations (dashed line).

Figure 4(b) extends Maeda's existing analysis by continuing to follow the trajectory of democratic states after they break down by depicting the likelihood that democracy will be restored. As indicated in Figure 2, we conceive of this process as consisting of three distinct transitions back to democracy, one for each type of democratic breakdown. That is, we allow the baseline hazard of democratic recovery to differ depending on how a democracy failed. Figure 4(b) indicates that states whose democratic breakdowns were exogenous (e.g., a coup) are least likely to experience a democratic restoration, as they have the highest probability of remaining in their current stage. The "other" category, however, appears to be quite short-lived. Democratic regimes that break down for other reasons have a low probability of persisting in that stage. Taken together, these survival curves begin to extend our understanding of democratic breakdowns and restorations, as those regimes that experience a failure of democracy *tend* not to remain in that stage for too long.

FIGURE 4. Non-Parametric Survival Plots – Maeda



(a) Democratic Breakdown (b) Democratic Recovery

Are these various baseline hazards different from one another? To adjudicate indirectly—and, admittedly, suboptimally—we compare the Akaike information criterion (AIC) goodness-of-fit statistic across two pairs of models (Table 2).[22] Table 2's two columns denote these pairings. We focus on transitions *out* of the Democracy stage in the first column, and transitions back *into* the Democracy stage in the second. Within each column, we estimate two models. We first estimate a basic non-parametric model in which we do not stratify the baseline hazard—we force all the transitions of interest to have the same baseline hazard ("No" row). We also estimate a model with stratified baseline hazards, which permits different baseline hazards for each transition of interest ("Yes" row, shaded). We then compare the AICs of the two models, within each column. Lower AICs represent models that better fit the data. If stratifying the baseline hazard is meaningful, we should see that the AIC for the shaded "Yes" row is lower than the "No" row's AIC. This is, in fact, exactly what we see. This suggests that stratifying the baseline hazard by transition type is meaningful, and produces results that better fit the data. The underlying rates at which different types of democratic breakdown, and democratic recovery occur are different.

[22] We can compare AICs for non-nested models, which will be the case here.

TABLE 2. AIC Comparisons - Maeda

DEMOCRATIC BREAKDOWN		DEMOCRATIC RESTORATION	
Stratified α_0(t)?	*AIC*	*Stratified* α_0(t)?	*AIC*
No	392.91	No	153.10
Yes	334.69	Yes	116.03

An exclusive focus on the risk of each individual transition, in isolation, is quite restrictive. It is poorly suited to assessing the larger democratic process, which comprises both the risk of democratic breakdown *and* the risk of democratic restoration following a breakdown. Substantively, we may be interested in the probability that a state is a democracy in five, ten or twenty years, but this simple query belies the fact that a state could be a democracy in five years either because it remained a democracy, *or* because it experienced a democratic breakdown and subsequently recovered within that five-year period. Transition probabilities provide exactly this type of inference, by generating an estimate of the probability that a state will occupy each stage in the model, taking into account every possible sequence of transitions through which a state may arrive in a given stage.

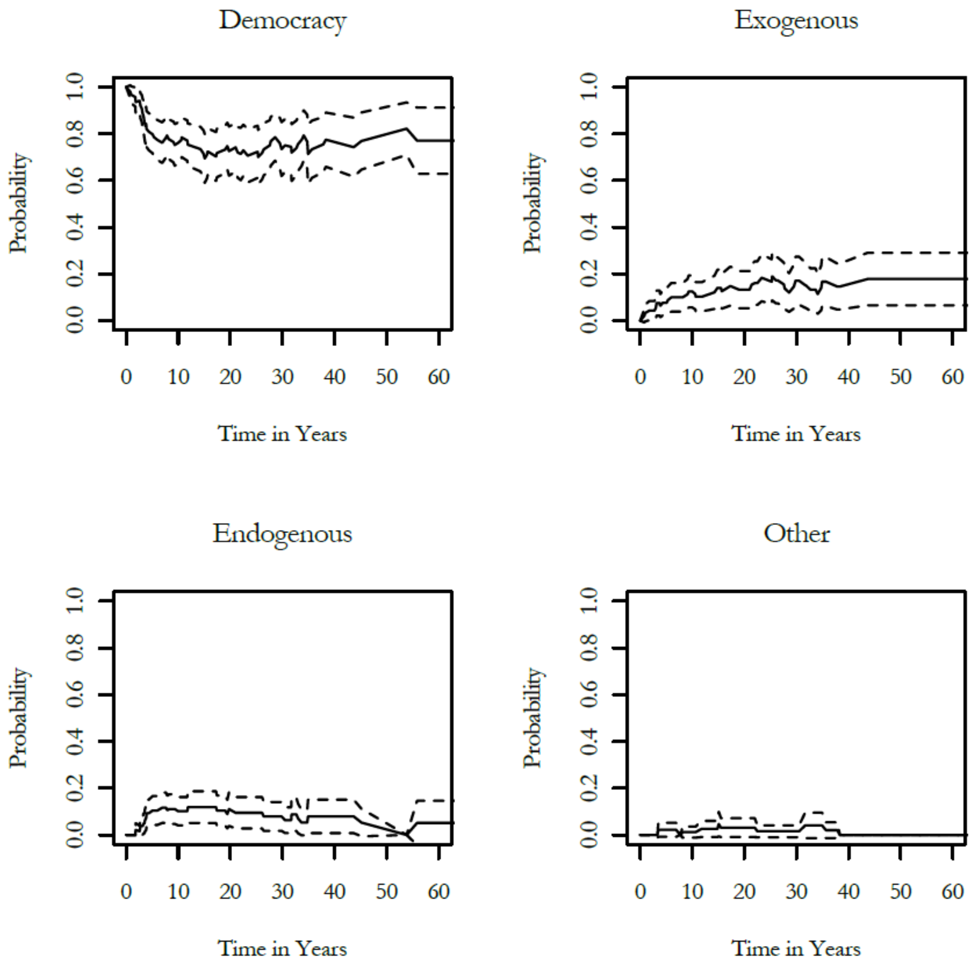
We begin by estimating a naïve set of transition probabilities by beginning with the observation time set to 0, and the state in the initial stage, Democracy. Figure 5 presents four plots, each presenting the probability that the state will occupy a different stage of the process over time, along with 95% confidence intervals. In essence, what this set of transition probabilities captures is: given that state *i* is a new democracy (t = 0), what is the probability that it will remain a democracy (Democracy, top left), experience a coup and remain undemocratic (Exogenous, top right), decide on its own to cease being a democracy (Endogenous, bottom left) or fail due to another reason (Other, bottom right).[23] These probability estimates take into account all of the different possible paths through which a state could arrive at each stage over time. For example, the probability that a state is democratic at time 10 reflects the probability that the state remained democratic over those 10 years, but also the probability that it

[23] Note that these probabilities sum to 1 because each state that enters the sample remains in the sample through the end of the observation period. Therefore, each state necessarily must occupy one stage of the process at any given point in time.

experienced an exogenous democratic failure and subsequently recovered, and the probability that it experienced an endogenous democratic failure and subsequently recovered, and so on.

Based on these naïve non-parametric estimates, it is clear that democratic states tend to remain democratic over time, and that after approximately 10 years, such states have roughly an 80% probability of being democratic (top left). Similarly, the other plots in Figure 5 indicate that the probability of a democratic state remaining undemocratic is quite low, though there is some evidence that exogenous forms of democratic breakdown may be stickier than other forms of breakdown.

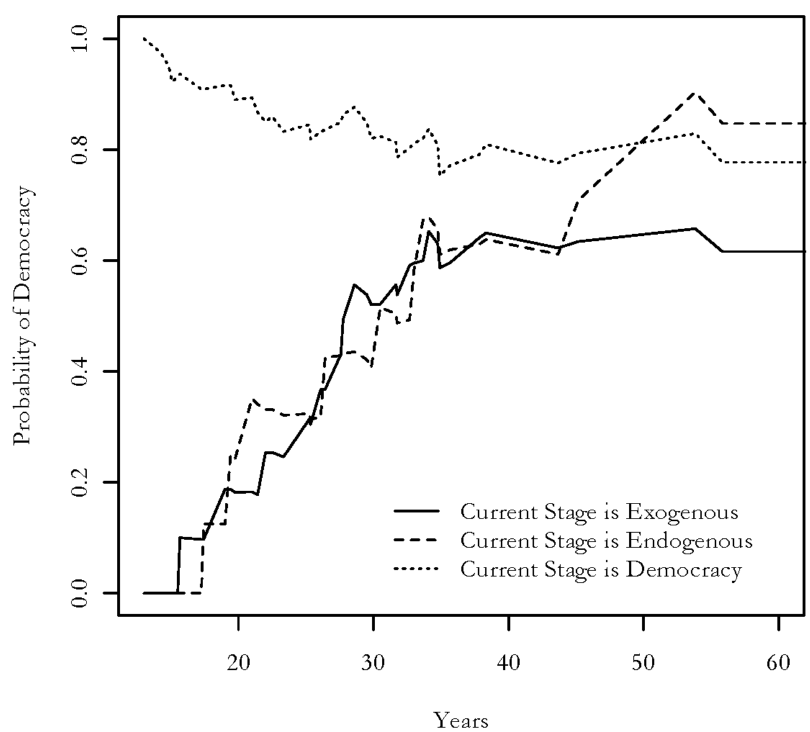
FIGURE 5. Naïve Non-Parametric Transition Probabilities



NOTE: Solid lines represent the probability of a state occupying the corresponding stage. Dashed lines represent 95% confidence intervals. All estimates begin with the current stage as Democracy, and time equal to 0.

In addition to these naïve estimates, it is also possible to estimate transition probabilities by varying both the stage that a state currently occupies, as well as the relevant time period. For example, it may be of substantive interest to consider the likelihood that a state will recover to democracy following a breakdown. To evaluate this scenario, Figure 6 plots the probability of a state occupying the Democracy stage, but varies the state's current stage. Moreover, the observation period begins at time 13, which is the average duration of the first democratic spell. Thus, Figure 6 plots the probability of a state being democratic in the future, given that 13 years after initially establishing a democracy, it is either in the Exogenous stage (solid line), the Endogenous stage (dashed line), or currently a Democracy (dotted line). Interestingly, Figure 6 indicates that the mode of democratic termination has relatively little impact on its likelihood of subsequent democratization. The probabilities of a state returning to Democracy, after either an exogenous or endogenous termination, are largely similar.

FIGURE 6. Non-Parametric Transition Probabilities of Democracy



NOTE: Estimates begin at the average duration of the first spell of democracy, 13 years. We vary the stage a state occupies at 13 years.

B. Huth and Allee: Territorial Dispute Escalation and Resolution[25]

Huth and Allee (2002) provide a foundational work on the escalation of territorial disputes. We re-analyze this particular dataset because Huth and Allee explicitly point to the importance of understanding territorial disputes as a long-term process comprising multiple sequential stages. Specifically, they "identify several stages or phases in an international dispute…[and contend] that any research design devised to test hypotheses about international conflict and cooperation should consider each of these possible stages" (2002, 22–23).

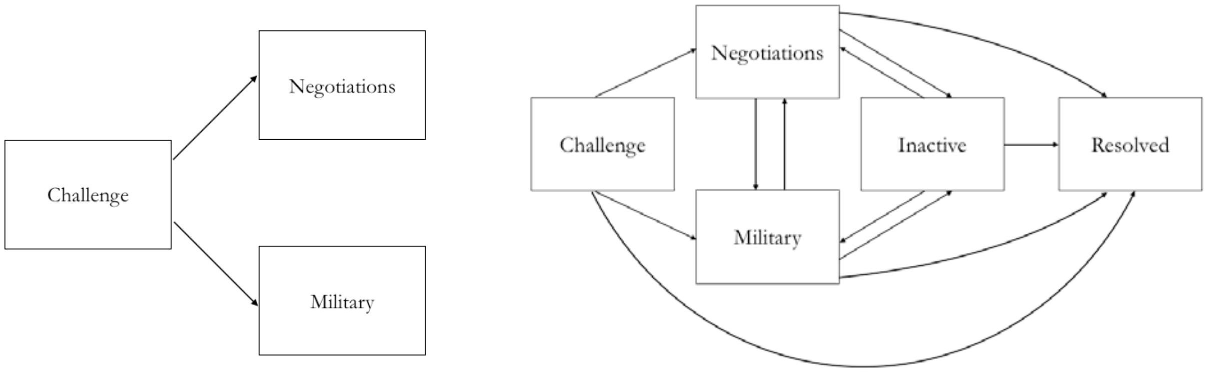
Huth and Allee's (2002) dataset contains 347 territorial disputes occurring anywhere in the world between 1919 and 1995 that involve independent states. The data are broken down further into *directed* state pairs over a dispute, where the states are ordered according to their relation to the status quo. States making explicit statements that challenge the territorial status quo (or statements challenging those made by another state) are listed first, followed by the state targeted by the challenger's statement (Huth and Allee 2002, 34). We refer to these dispute-challenger-target combinations as "dispute-dyads." There are 398 dispute-dyads in Huth and Allee's dataset.

Huth and Allee's substantive interest is in the different mechanisms undergirding the democratic peace, and how the mechanism of interest may differ depending on the resolution method.[27] To investigate this, Huth and Allee use a multinomial logit model, in which they set "no resolution method chosen" (what they call "Challenge") as the reference outcome. Their model choice allows them to parse out the effect of their variables of interest on the probability of peaceful negotiations vs. "no resolution", and the probability of militarized behavior vs. "no resolution" (2002, chap. 7).

[25] This section liberally borrows from another paper of ours, for the time being (Jones and Metzger 2015). In the future, we plan to include different applications.

[27] They explicate three major theoretical models related to the democratic peace: the Political Accountability model, the Political Norms model, and the Political Affinity model (Huth and Allee 2002, 67).

FIGURE 7. The Territorial Dispute Process



(a) Limited Dispute Process (b) Complete Dispute Process

Nevertheless, this research design is not ideal for fully examining the dispute, as a process. It artificially restricts the manner in which a dispute might evolve over time. With their multinomial logit models, Huth and Allee are able to examine the initial transition from a challenge to either formal negotiations *or* militarized behavior (Figure 7(a)). However, this setup is incapable of capturing the inherently recursive nature of many disputes: a dispute might first militarize, revert to an inactive period,[28] only to then to transition to negotiations. Moreover, this analysis necessarily omits a fundamental outcome of interest in the study of territorial disputes: their ultimate resolution. Huth and Allee's analysis is only capable of modeling either (1) the first transition within the dispute process, thus ignoring all subsequent events that may occur; or (2) truncating the analysis to only include transitions into Negotiations or Military, with no analysis of what happens after a dispute-dyad transitions into one of those two stages.

Our multi-state modeling strategy bypasses many of these weaknesses, by modeling the complete process as depicted in Figure 7(b). In this approach, we conceive of 5 possible stages: the original three identified by Huth and Allee, but also periods of inactivity,[28] as well as a fifth, absorbing stage, indicating that a challenge to the territorial status quo has been resolved.[29] Table 3 provides information on the observed transitions in Huth and Allee's expanded dataset.

[28] In inactive periods, the dispute is ongoing, but the two states are not actively attempting to resolve the dispute.

[29] Huth and Allee do allow for the possibility of periods of inactivity, but they treat these periods as being the same as the initial challenge phase and include a counter for the number of prior active settlement attempts (militarizations

TABLE 3. Observed Transition Frequencies – Huth & Allee

Current Stage	*Next Stage* Negotiation	Military	Inactive	Resolved	TOTAL (for %s)
Challenge	278	71	–	7	356
(Row Total %)	(77.7%)	(19.8%)		(2.0%)	(99.5%)
Negotiations	–	16	1514	206	1736
(Row Total %)		(0.9%)	(86.6%)	(11.8%)	(99.3%)
Military	13	–	315	55	383
(Row Total %)	(3.4%)		(81.6%)	(14.2%)	(99.2%)
Inactive	1427	286	–	67	1780
(Row Total %)	(78.0%)	(15.6%)		(3.7%)	(97.3%)
TOTAL	1718	373	1829	335	4255

"TOTAL" column contains the total number of transitions across the row, representing the number of transitions from the current stage to another stage. The percentage calculations use the number of overall dispute-dyads in the current stage as the denominator. They do not sum to 100% because of right censoring. Cells containing a dash indicate impossible transitions in our model.

[Insert Table 4 here]

We re-estimate the territorial dispute process using a multi-state model to highlight the additional implications that this modeling approach reveals. As with the previous example, we stratify the baseline hazard for each transition depicted in Figure 7(b). We present the results of the multi-state model in Table 4, using the first initial of each stage to denote transitions.

1. SEMI-PARAMETRIC

We begin by performing a specification test. After all, it may not even be necessary to estimate transition-specific covariates. To assess this, we use a likelihood-ratio test (Aalen, Borgan, and Gjessing 2008, 135–136). We take the model we report in Table 4, and compare it to a second model in which all the covariates are forced to have the same effect across every transition. The second model is a restricted version of Table 4's model, since we are constraining the parameter estimates to be equal across transitions. The null hypothesis is that Table 4's coefficients and the second model's coefficients are

or negotiations). We estimate both a 4- and 5-stage model, with the former treating all periods of inactivity as pertaining to the Challenge stage, and the latter differentiating between periods of inactivity following a settlement attempt and the initial challenge. We present the 5-stage model above as it has a lower AIC and BIC score, indicating that although it includes additional parameters, a model with 12 strata provides a better fit to the data.

equal. A significant likelihood-ratio test means that the coefficients are *not* equal, implying that our unconstrained model from Table 4 is the better bet. Our test comes back with a *p*-value less than 0.05 ($p = 0.000$, $\chi^2 = 252.11$ with 66 d.f.). Therefore, our use of transition-specific covariates appears to be justified.

Broadly speaking, our results mirror those of Huth and Allee with respect to the determinants of whether a dispute-dyad experiences negotiations or militarization. Democratic challengers increase the likelihood of formal negotiations (C → N), and reduce the likelihood of militarization after a challenge has been initiated (C → M). Moreover, this pacifying effect of challenger regime type is also observed in subsequent transitions. Following periods of inactivity, dispute-dyads with democratic challengers are still more likely to experience negotiations (I → N) and less likely to experience militarization (I → M). This finding is consistent with the expectations of the democratic peace theory by demonstrating that democracies tend to favor non-violent forms of dispute resolution.

However, the use of a multi-state model advances our understanding of the role of regime type in territorial disputes by revealing a greater degree of complexity than previously noted. In this instance, we find a somewhat surprising result for the effect of challenger regime type on the likelihood of dispute resolution. When a dispute-dyad is currently engaged in negotiations, the challenger's regime type has no statistically significant effect on the likelihood of resolving the dispute (N → R). However, if a dispute-dyad is currently engaged in militarization, democratic challengers are significantly less likely to resolve the dispute than their autocratic counterparts (M → R), and are more likely to remain in the militarization stage for longer. This result indicates that democratic challengers, though less likely to become involved in militarizations in the first place, will typically remain in periods of militarization for *longer*, delaying the ultimate resolution of the dispute.

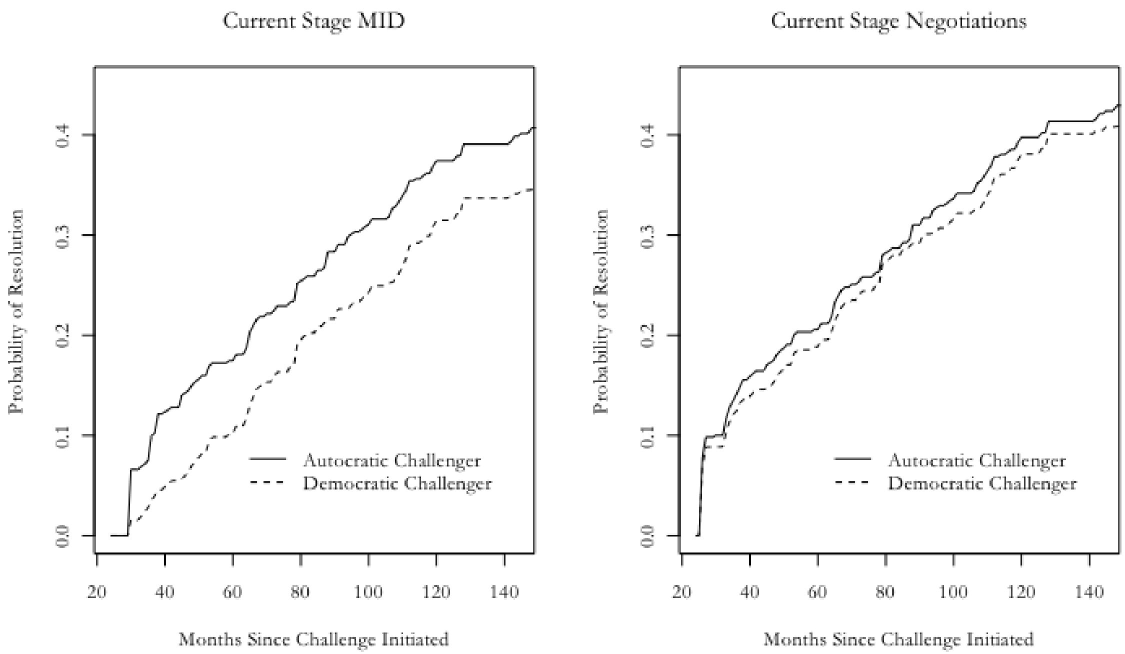
We prod this result further by assessing whether the two individual coefficients are significantly different from one another, using a Wald test. The null hypothesis for a Wald test of this sort is that the coefficients are equal ($\beta_{(N \rightarrow R)DEM} = \beta_{(M \rightarrow R)DEM}$). Our Wald test is statistically significant ($p = 0.047$, $\chi^2 = 3.95$ with 1 d.f.), suggesting that the coefficients are indeed statistically different from one another.

This counterintuitive finding is quite interesting as it calls into question the straightforward democratic peace conclusion that democratic challengers favor peaceful forms of dispute resolution. By taking into account the entire process of territorial disputes, rather than focusing exclusively on the initial phase, it is possible to provide a more comprehensive assessment of the role of democracy in territorial disputes. In order to determine what the overall effect of democracy is on the territorial dispute process, taking into account *both* that disputes involving democratic challengers are less likely to militarize *and* that disputes involving democratic challengers that do militarize will tend to last longer, we estimate a series of transition probabilities for challenger regime type. Figure 8 compares the probability that a dispute will occupy the Resolved stage for challengers with a Polity score of +7 (75th percentile, dashed line) vs. challengers with a Polity score of -8 (25th percentile, solid line). The panels in Figure 8 begin two years after the initial challenge, for two different 'stage' scenarios: the dispute is in the Military stage (left panel), vs. the Negotiations stage (right panel).[30]

Figure 8 shows that, when a dispute-dyad is engaged in formal negotiations after two years (right panel), the challenger's regime type has no meaningful impact on the probability of the dispute being subsequently resolved. By contrast, when a dispute-dyad is engaged in a MID after two years (left panel), the challenger's regime type plays a more appreciable role. The dispute is less likely to be resolved when the challenger is a democracy, compared to when the challenger is an autocracy. Nevertheless, this indicates territorial disputes will subsequently stay unresolved longer for democratic challengers than autocratic ones after a militarization, whereas no such difference exists following formal talks.

[30] All other covariates are held at their median values. Two years is the average duration that dispute-dyads remain in the initial challenge stage.

FIGURE 8. The Effect of Challenger Regime Type on Dispute Resolution



NOTE: Each panel depicts the probability of a dispute-dyad being Resolved at various points in time, given that the dispute has transitioned to either a MID (left panel) or Negotiations (right panel) after 24 months.

IV. Conclusion

How can we model durations in processes comprised of multiple stages? We suggest that multi-state survival models are one answer. Estimated as stratified Cox models, the model permits researchers to examine all of the transitions that constitute a process. Multi-state models are incredibly flexible, and are able to capture many different stage structures. They can easily accommodate processes with competing transitions, recursive transitions, and repeated transitions. The end result is a more holistic take on the process of interest. Simpler survival models, like the standard Cox model and competing risks, are less holistic, as they are more restrictive. Standard Cox models only examine a single *transition*, and competing risks models only examine a single *stage* (and every transition out of that stage).

We highlighted three features that make multi-state models particularly attractive: transition-specific baseline hazards, transition-specific covariate effects, and overall transition probabilities. The first two make the last possible. The last is also particularly important, and epitomizes multi-state models' holistic perspective. The models can estimate the probability of a subject occupying a particular stage at *t* by accumulating the probabilities associated with every possible transition sequence—both direct and indirect—that ends in that stage.

We used two different applications to showcase these features. We first extended Maeda's (2010) study of democratic reversion by adding democratic restorations. We used a non-parametric specification to highlight the different baseline hazards, and how to assess their equivalence. In our application, we saw that the baseline hazards out of Democracy were not significantly different from one another, nor were the baseline hazards back into Democracy. Yet, stratifying the hazards did improve model fit, as we demonstrated with AIC. We also estimated overall transition probabilities into Democracy from the non-parametric specification, which (unsurprisingly) bore out more of the same: when we varied the state's current stage, few appreciable differences existed across the transition probabilities.

Our second application looked at Huth and Allee's (2002) study of the democratic peace and territorial disputes. We extended their study, and examined more than just transitions out of the initial Challenge stage. Here, we used a semi-parametric specification to bring out the importance of transition-specific covariate effects, and again, how to assess their equivalence. We employed a likelihood-ratio test to test whether we needed transition-specific covariates at all; the test came back in the affirmative. We found that the effect of democracy varies depending on the transition in question, in a way that Huth and Allee's initial study did not uncover. We also talked about the differing effects of democracy on the probability of resolution, from the Military (M → R) and the Negotiation stages (N → R). A Wald test of the two corresponding coefficients shows them to be significantly different from one another. We then estimated overall transition probabilities, to see the net effect of democracy on resolving the dispute. We varied the current stage, and computed transition probabilities for democratic and non-democratic challengers. We found that democratic challengers' probability of transitioning into Resolved is indistinguishable from non-democracies when the current stage is Negotiations, but there is a more appreciable difference between the two regime types when the current stage is Militarization.

Works Cited

Aalen, Odd, Ornulf Borgan, and Hakon Gjessing. 2008. *Survival and Event History Analysis: A Process Point of View*. New York: Springer.

Allison, Paul D. 1984. *Event History Analysis: Regression for Longitudinal Event Data*. Thousand Oaks, CA: Sage.

Beck, Nathaniel, Jonathan Katz, and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42 (4): 1260–1288.

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.

Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14 (1): 63–82.

Chiba, Daina, Lanny W. Martin, and Randolph T. Stevenson. 2015. "A Copula Approach to the Problem of Selection Bias in Models of Government Survival." *Political Analysis* 23 (1): 42–58.

Chiba, Daina, Nils W. Metternich, and Michael D. Ward. Forthcoming. "Every Story Has a Beginning, Middle, and an End (But Not Always in That Order): Predicting Duration Dynamics in a Unified Framework." *Political Science Research and Methods*.

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

Diehl, Paul F. 2006. "Just a Phase?: Integrating Conflict Dynamics Over Time." *Conflict Management and Peace Science* 23 (3): 199–210.

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

Fine, Jason P., and Robert J. Gray. 1999. "A Proportional Hazards Model for the Subdistribution of a Competing Risk." *Journal of the American Statistical Association* 94 (446): 496–509.

Fukumoto, Kentaro. 2009. "Systematically Dependent Competing Risks and Strategic Retirement." *American Journal of Political Science* 53 (3): 740–754.

Gordon, Sanford C. 2002. "Stochastic Dependence in Competing Risks." *American Journal of Political Science* 46 (1): 200–217.

Greene, William H. 2012. *Econometric Analysis*. 7th ed. Upper Saddle River, NJ: Prentice Hall.

Hays, Jude C., and Aya Kachi. 2009. "Interdependent Duration Models in Political Science." Presented at the Annual Meeting of the American Political Science Association, Toronto, ON.

Huth, Paul K., and Todd L. Allee. 2002. *The Democratic Peace and Territorial Conflict in the Twentieth Century*. Cambridge: Cambridge University Press.

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

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

Jones, Benjamin T., and Eleonora Mattiacci. 2015. "(Nuclear) Change of Plans: What Explains Nuclear Reversals?" Presented at the Annual Meeting of the International Studies Association, New Orleans, LA.

Jones, Benjamin T., and Shawna K. Metzger. 2015. "Evaluating Conflict Dynamics: Many Conceptualizations, A Novel Empirical Approach." Presented at the Annual Meeting of the International Studies Association, New Orleans, LA.

King, Gary, James E. Alt, Nancy Elizabeth Burns, and Michael Laver. 1990. "A Unified Model of Cabinet Dissolution in Parliamentary Democracies." *American Journal of Political Science* 34 (3): 846–871.

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

Maeda, Ko. 2010. "Two Modes of Democratic Breakdown: A Competing Risks Analysis of Democratic Durability." *Journal of Politics* 72 (4): 1129–1143.

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

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.

Prentice, R. L., B. J. Williams, and A. V. Peterson. 1981. "On the Regression Analysis of Multivariate Failure Time Data." *Biometrika* 68 (2): 373–379.

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.

Skrondal, Anders, and Sophia Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. London: Chapman & Hall/CRC.

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

Therneau, Terry M., and Patricia M. Grambsch. 2000. *Modeling Survival Data: Extending the Cox Model*. New York: Springer.

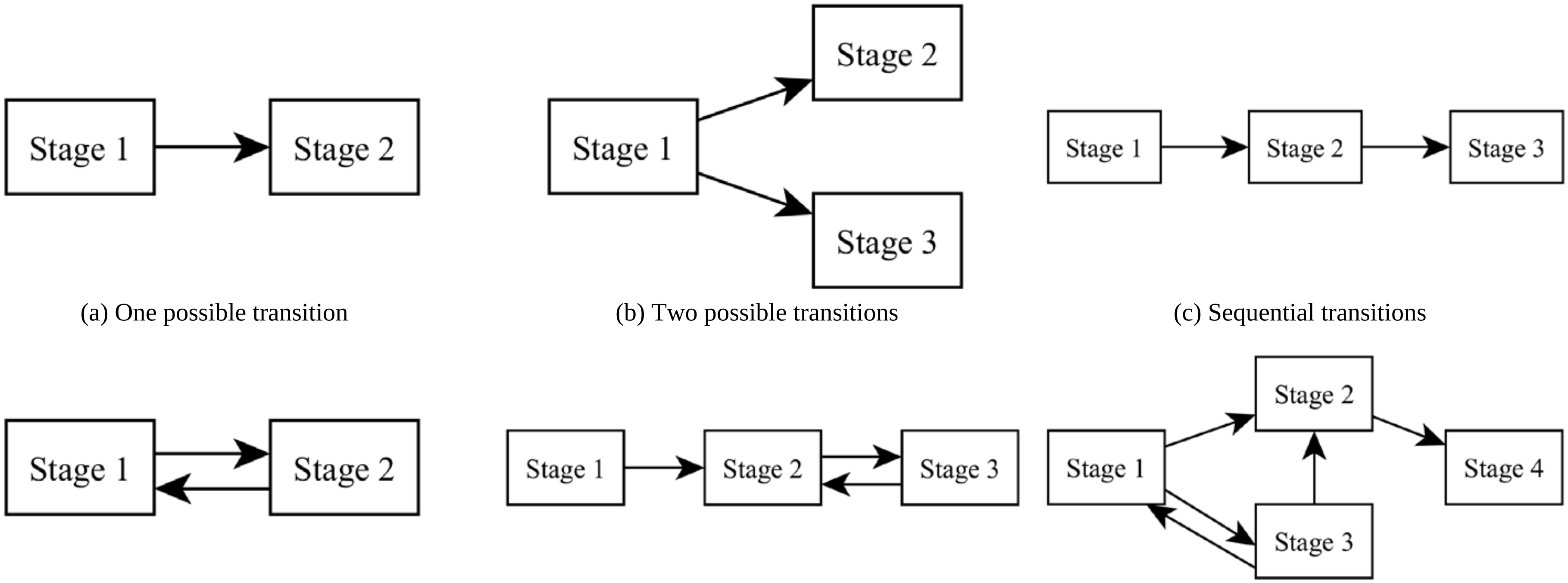
Vollset, Stein Emil, Aage Tverdal, and Håkon K. Gjessing. 2006. "Smoking and Deaths between 40 and 70 Years of Age in Women and Men." *Annals of Internal Medicine* 144 (6): 381–389.

Weskamp, Paul, and Markus Höchstötter. 2010. "Change Point Analysis and Regime Switching Models." Working paper. https://statistik.econ.kit.edu/download/Change_Point_Analysis.pdf.

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.

------. 2011. "mstate: An R Package for the Analysis of Competing Risks and Multi-State Models." *Journal of Statistical Software* 38 (7): 1–30.

FIGURE 1. Illustrative Processes



(a) One possible transition (b) Two possible transitions (c) Sequential transitions

(d) Repeated transitions (e) Both sequential and repeated transitions (f) All previous panels

Arrows denote possible transitions for each example process.

TABLE 4. Multi-state Model of the Territorial Dispute Process – 5 Stages

	C → N	C → M	C → R	N → I	N → M	N → R
Ratio of Military Capabilities	-0.152	0.471	-0.912	-0.295**	0.746	0.702*
	(0.248)	(0.473)	(2.404)	(0.113)	(1.128)	(0.297)
Strategic Value	0.268†	0.757**	-0.702	0.128*	0.337	-0.010
	(0.141)	(0.253)	(1.229)	(0.062)	(0.590)	(0.170)
Target Engaged in Other Dispute	-0.189	0.954***	0.735	-0.051	1.417*	0.168
	(0.147)	(0.258)	(1.000)	(0.065)	(0.618)	(0.176)
Challenger Engaged in Other Dispute	0.235	0.573*	1.011	0.072	0.292	0.318†
	(0.153)	(0.269)	(1.180)	(0.072)	(0.689)	(0.178)
Challenger Regime Type	0.032***	-0.033†	-0.022	0.004	0.002	-0.006
	(0.009)	(0.019)	(0.067)	(0.004)	(0.038)	(0.010)
Target Regime Type	-0.003	-0.047*	0.156	-0.006	-0.033	0.001
	(0.010)	(0.019)	(0.114)	(0.004)	(0.040)	(0.010)

	M → I	M → N	M → R	I → N	I → M	I → R
Ratio of Military Capabilities	-0.090	2.466	1.472*	-0.124	1.569***	0.576
	(0.290)	(1.718)	(0.734)	(0.114)	(0.241)	(0.526)
Strategic Value	-0.316*	-0.058	0.468	0.132*	0.098	-0.097
	(0.159)	(0.777)	(0.364)	(0.060)	(0.133)	(0.294)
Target Engaged in Other Dispute	-0.313*	-0.516	-0.268	-0.059	0.189	0.400
	(0.144)	(0.883)	(0.364)	(0.063)	(0.140)	(0.278)
Challenger Engaged in Other Dispute	-0.058	-0.395	0.219	-0.087	0.224	0.569*
	(0.157)	(0.743)	(0.365)	(0.070)	(0.139)	(0.278)
Challenger Regime Type	0.011	0.038	-0.063*	0.016***	-0.055***	0.013
	(0.011)	(0.050)	(0.026)	(0.004)	(0.010)	(0.018)
Target Regime Type	0.028**	0.076	-0.031	-0.002	0.012	0.013
	(0.010)	(0.053)	(0.023)	(0.004)	(0.008)	(0.017)
Log-Likelihood (partial)	-15601.52					

† = $p \leq 0.10$, * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$, two-tailed tests.

NOTE: C = Challenge; N = Negotiations; M = Military; R = Resolved; I = Inactive

Appendix A
Dataset Structure: The Details

We create our analysis dataset by reshaping Huth and Allee's "status quo" dataset (2002, chap. 7). Huth and Allee's dataset contains information on "whether, when, and how states with territorial claims initiate foreign policy actions intended to alter the territorial status quo" (Huth and Allee 2009, 1). Table 4 shows the first five observations from one territorial dispute in the dataset (over the Maynas region), to give a sense of the dataset's structure, and the subsequent transformations we perform.

TABLE 4. H&A: ECU-PER, Maynas Region, 1/1950-8/1953

Challenger	*Target*	*Stage Start Date*	*Stage*	*Time in Stage (incl.)*	*Time Since Last Action (incl.)*
Ecuador	Peru	9/1950	Negotiations	2	9
Ecuador	Peru	8/1951	Negotiations	1	10
Ecuador	Peru	9/1951	Challenge	12	1
Ecuador	Peru	9/1952	Negotiations	1	13
Ecuador	Peru	2/1953	Military	7	5

Dispute begins in 1950 (Huth and Allee 2002, 447). All times are measured in months.
Incl. = includes start date month.

Huth and Allee make two coding decisions that are notable because they have ramifications down the line for our efforts. First, Huth and Allee permit only one ongoing settlement attempt at a time. For example, if a challenger initiates negotiations, and then also initiates a MID while negotiations are ongoing, only the negotiation attempt enters Huth and Allee's dataset, because it started first.[31] We decided to leave Huth and Allee's data, as is, instead of trying to add information on simultaneous settlement attempts. Our primary purpose is replication, and simultaneous settlement attempts were simply less relevant for Huth and Allee's original study.

[31] We verified this by comparing Huth and Allee's dataset with the territorial dispute data from the Issue Correlates of War (ICOW) project (Hensel 2001), as we knew that ICOW's coding rules would record any simultaneous ongoing settlement attempts. We took care to distinguish between challenger-initiated settlement attempts (which are Huth and Allee's focus) with *any* settlement attempt over the issue (which is what ICOW codes). We used the MID data to recover information on which state initiated the militarized settlement attempt in ICOW. Initiation information is not readily available for peaceful settlement attempts. Once we identified a few potential cases with simultaneous settlement attempts, we searched the internet on a case-by-case basis to find who initiated the peaceful settlement attempt.

Second, Huth and Allee adopt specific coding rules to deal with the Challenge stage. The authors use a "twelve-month rule" to determine "when" to add observations for the Challenge stage. Specifically, they add an observation for the Challenge stage at the start of every 12-month span in a dispute where the challenger does not (a) initiate negotiations with the target or (b) take military action toward the target (2002, 142). For example, between September 1951 and August 1952 (inclusive), Ecuador took no action to resolve its dispute with Peru over the Maynas region. As a result, Huth and Allee add one Challenge stage observation in 9/1951 (Table 4, 3rd observation) for the ECU-PER Maynas dispute.

TABLE 5. ECU-PER, Maynas Region, 1950 only
Dispute-Dyad-Month, Discrete

Challenger	*Target*	*Date*	*Stage*
Ecuador	Peru	1/1950	Challenge
Ecuador	Peru	2/1950	Challenge
Ecuador	Peru	3/1950	Challenge
Ecuador	Peru	4/1950	Challenge
Ecuador	Peru	5/1950	Challenge
Ecuador	Peru	6/1950	Challenge
Ecuador	Peru	7/1950	Challenge
Ecuador	Peru	8/1950	Challenge
Ecuador	Peru	9/1950	Negotiations
Ecuador	Peru	10/1950	Negotiations
Ecuador	Peru	11/1950	Challenge
Ecuador	Peru	12/1950	Challenge

Our reshaping efforts can be broken into three steps. We begin by converting Huth and Allee's data into a pure dispute-dyad-month structure. Our goal was to ensure that one observation existed for every month that the dispute-dyad was active.[32][33] This meant "filling in the gaps" between the original

[32] Our reasons are twofold. First, we are interested in modeling a dispute-dyad's transitions between various stages. This means we need detailed information on which stage a dispute-dyad is in (and when), so that we can identify which transitions are possible (and when, again). Second, we are also interested in how the probability of certain transitions can change as a dispute-dyad spends more time in a stage. To calculate time-in-stage for a dispute-dyad, we need the stage's start date and its end date. Huth and Allee's dataset structure, in its original form, can provide incorrect information regarding both points. (E.g., fn. 33.)

[33] We corrected the start and end dates for "colonial legacy" disputes (e.g., fn. 18) by hand. We coded dispute-dyads containing newly independent colonies as beginning in the same month as the former colony's independence. We

dispute-dyad observations. Sometimes, it also meant adding observations to the beginning of a dispute-dyad. This happened any time a dispute-dyad experienced a negotiation or military action within 12 months of its initiation, because of the twelve-month rule. Table 4's dispute is one example: Ecuador and Peru entered into negotiations over the Maynas in September 1950, 9 months after the dispute began. Table 5 displays the result of this conversion for Ecuador and Peru's Maynas dispute, for 1950 only (to save space).

Second, we converted the dataset back into a continuous-time format, which is what the `mstate` package requires (Wreede, Fiocco, and Putter 2010, 2011). Every row in this intermediate dataset represents a unique combination of three factors: dispute, dyad, and stage start date. Table 6 shows the resultant observations for Ecuador and Peru's Maynas dispute, for the same temporal range as Table 4. Notice how Table 6 has more observations compared to Table 4. The difference stems from the Challenge observations and the twelve-month rule. Ecuador and Peru's dispute is in the Challenge stage on four separate occasions (Table 6). However, only one of the four Challenge-stage visits is for 12 months or longer (the visit that starts in 9/1951), making this the only observation that qualifies for inclusion under the twelve-month rule in Table 4.

TABLE 6. ECU-PER, Maynas Region, 1/1950-8/1953
Dispute-Dyad-Stage Start Date, Continuous

Challenger	*Target*	*Stage St. Date*	*Stage (Current)*	*Next Stage*	*Time in Stage (incl.)*
Ecuador	Peru	1/1950	Challenge	Negotiations	8
Ecuador	Peru	9/1950	Negotiations	Challenge	2
Ecuador	Peru	11/1950	Challenge	Negotiations	9
Ecuador	Peru	8/1951	Negotiations	Challenge	1
Ecuador	Peru	9/1951	Challenge	Negotiations	12
Ecuador	Peru	9/1952	Negotiations	Challenge	1
Ecuador	Peru	10/1952	Challenge	Military	4
Ecuador	Peru	2/1953	Military	Challenge	7

All times are measured in months. Incl. = includes start date month.

then coded the preceding dispute-dyad, involving the former colonial power, as ending the month prior. For independence dates, we used the system-entry dates from the Correlates of War project's State Membership list.

Third and finally, we add information regarding *all* the possible transitions associated with a particular stage. Why we need this information is best understood using the example. The first observed transition in Ecuador and Peru's dispute is Challenge → Negotiations. However, this is not the only possible transition out of the Challenge stage, as Figure 1c makes clear. We could have also seen Challenge → Military, or Challenge → Resolved. Neither possibility is acknowledged in Table 6's intermediate dataset. Given our (and multi-state models') interest in these other transitions, the omission is troublesome. Similarly for the second observation, there are three possible transitions from the Negotiations stage: Negotiations → Challenge, Negotiations → Military, and Negotiations → Resolved. Again, only the observed transition, Negotiations → Challenge, appears in the dataset, with no hint as to the other two possible transitions.

To rectify this, we create duplicate observations for every possible transition out of a stage (Wreede, Fiocco, and Putter 2010, 263). Doing so effectively triples the number of observations in our dataset, since each of our stages has three possible exiting transitions (with the exception of the Resolved stage). The result is the final dataset that we use in our analysis. Every row in our dataset represents a unique combination of *four* factors: dispute, dyad, stage start date, and possible transition.[34] Table 7 shows how our running example of ECU-PER appears in our final dataset, for the same temporal range as Table 4. With this dataset structure, we can now use `mstate` in `R` to estimate the stratified Cox model that underlies our multi-state model (Wreede, Fiocco, and Putter 2011).[35]

[34] There are nine possible transitions in our model (Figure 1c). We assign arbitrary numerical identifiers to each transition, for ease of reference. The identifier is denoted as q in Equations 1 and 2.

[35] `mstate` also requires that each independent variable appears q times—that is, one copy of the variable for every possible transition. For further coding details, see Wreede, Fiocco, and Putter (2010, 263).

TABLE 7. Final Dataset Structure: ECU-PER, Maynas Region, 1/1950-8/1953
Dispute-Dyad-Stage Start Date-q, Continuous

Challenger	*Target*	*Stage St. Date*	*Stage (Current)*	*Next Possible Stage*	*Transition Occurs?*	*Time in Stage (incl.)*	*Transition ID (q)*
Ecuador	Peru	1/1950	Challenge	Negotiations	Yes	8	1
Ecuador	Peru	1/1950	Challenge	Military	No	8	2
Ecuador	Peru	1/1950	Challenge	Resolved	No	8	3
Ecuador	Peru	9/1950	Negotiations	Challenge	Yes	2	4
Ecuador	Peru	9/1950	Negotiations	Military	No	2	5
Ecuador	Peru	9/1950	Negotiations	Resolved	No	2	6
Ecuador	Peru	11/1950	Challenge	Negotiations	Yes	9	1
Ecuador	Peru	11/1950	Challenge	Military	No	9	2
Ecuador	Peru	11/1950	Challenge	Resolved	No	9	3
Ecuador	Peru	8/1951	Negotiations	Challenge	Yes	1	4
Ecuador	Peru	8/1951	Negotiations	Military	No	1	5
Ecuador	Peru	8/1951	Negotiations	Resolved	No	1	6
Ecuador	Peru	9/1951	Challenge	Negotiations	Yes	12	1
Ecuador	Peru	9/1951	Challenge	Military	No	12	2
Ecuador	Peru	9/1951	Challenge	Resolved	No	12	3
Ecuador	Peru	9/1952	Negotiations	Challenge	Yes	1	4
Ecuador	Peru	9/1952	Negotiations	Military	No	1	5
Ecuador	Peru	9/1952	Negotiations	Resolved	No	1	6
Ecuador	Peru	10/1952	Challenge	Negotiations	No	4	1
Ecuador	Peru	10/1952	Challenge	Military	Yes	4	2
Ecuador	Peru	10/1952	Challenge	Resolved	No	4	3
Ecuador	Peru	2/1953	Military	Challenge	Yes	7	7
Ecuador	Peru	2/1953	Military	Negotiations	No	7	8
Ecuador	Peru	21953	Military	Resolved	No	7	9

All times are measured in months. Incl. = includes start date month.