# What Happens Depends on When It Happens: Copula-based Ordered Event History Analysis of Civil War Duration and Outcome *

Kentaro Fukumoto †

*This is the submitted, author-prepared version of the manuscript which is accepted by *Journal of American Statistical Association* (http://www.tandfonline.com/doi/abs/10.1080/01621459.2014.979994#.VN2BbvmG8dc). This work was supported by the Japan Society for the Promotion of Science (grant number 20330023) and Abe Yoshishige Memorial Educational Fund. The author appreciates David E. Cunningham, Kristian Skrede Gleditsch, and Idean Salehyan for making their data available and Janet Box-Steffensmeier, Frederick Boehmke, Robert Franzese, Jude Hays, John Londregan and Yasuhiro Omori for their comments.

†Kentaro Fukumoto is Professor, Department of Political Science, Faculty of Law, Gakushuin University, 1-5-1 Mejiro, Toshima, Tokyo, 171-8588, Japan (E-mail: *Kentaro.Fukumoto@gakushuin.ac.jp*) and Visiting Fellow, Department of Political Science, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO 63130-4899, U.S.A.

**Abstract**

Scholars are interested in not just what event happens but also when the event happens. If there is dependence among events or dependence between time and events, however, the currently common methods (e.g. competing risks approaches) produce biased estimates. To deal with these problems, this article proposes a new method of copula-based, ordered event history analysis (COEHA). A merit of working with copulas is that, whatever marginal distributions time and event variables follow (including the Cox model), researchers can derive whatever joint distribution exists between the two. Application of the COEHA model to a dataset of civil war answers two controversial questions. First, as wars become longer, rebel victory becomes more likely but settlement does not. Second, stronger rebels make wars longer but do not necessarily tend to win, as experts predict but fail to establish.

Key Words: Competing risk, Interdependence, Separate estimation, Rebel, Intrastate conflict

Note: Appendices and replication files for this article are available online.

# 1. INTRODUCTION

Scholars have been interested in not just what event happens but also when those events happen. The running example of this article is civil war duration and outcome (e.g. government victory, rebel victory or settlement). At first, scholars examined these two aspects *separately*, which I call a "separate estimation" approach. Some studies analyze duration, but not outcome, by event history analysis (Regan 2002; Fearon 2004). Other works consider outcomes, but not duration, by (multinomial) logit models (Mason and Fett 1996; Mason et al. 1999), where duration is sometimes an *exogenously* given *independent* variable.

Recently, social science research is beginning to deal more directly with the fact that what happens depends on when it happens, because people behave strategically. In the context of civil war, how or why a conflict ends is related to when it ends, because "[g]overnment and rebels make decisions with both the outcome and expected duration in mind" (DeRouen and Sobek 2004, p. 304). In order to deal with both time and events *simultaneously*, researchers employ a "competing risks" (hereafter, CR) approach (DeRouen and Sobek 2004; Balch-Lindsay et al. 2008; Brandt et al. 2008; Cunningham et al. 2009; Thyne 2009; Akcinaroglu 2012). Here, duration is an *endogenously* generated *dependent* variable.

This article, however, argues that even the CR model has two problems. First, the CR model's assumption of conditional independence among risks is violated if the timing of one type of event's occurrence depends on the potential timing of another type, even after controlling for some covariates. I call the situation "dependence among events." In general, categories of many discrete event variables are of an ordered nature and, in that case, there should be dependence among events. Second, the CR model suffers from bias if there still remain omitted variables (e.g. unobserved or unobservable military strength of government or rebels) which affect both duration and outcome even after controlling for some observed covariates. I call the situation "dependence between time and events." Unfortunately, in the real world, these two kinds of dependence often arise. In an example of civil war, Cunningham et al. (2009, p. 587) are concerned that, "if one of the parties is likely to win a conflict militarily, the willingness of the other party to make concessions should also increase and make agreements more likely. Furthermore, certain outcomes may become or less likely over time relative to the others." That is, they warn that dependence among events and dependence between time and events are present in their civil war data.

To address these two types of dependence, this article proposes a new method: copula-based ordered event history analysis (COEHA). It models the joint distribution of latent time and ordered event variables by taking advantage of a "copula" function (which will be explained shortly). A merit of the copula is its flexibility and modularity, that is, whatever marginal distributions time and event variables follow, analysts can derive any joint distribution that exists between the two. This study applies the COEHA model to the civil war dataset of Cunningham et al. (2009) and gives answers to two controversial questions. First, some studies demonstrate that longer wars lead to settlement, while other works show that longer wars result in rebel victory. This article agrees with the latter finding. Second, Cunningham et al. (2009) predict that stronger rebels do not necessarily win, though their own analysis using the CR model implies that stronger rebels tend to prevail. My reanalysis of their data with the COEHA model supports the authors' original prediction.

Besides intrastate conflict duration and outcome, social science is abundant in contexts to which we can apply the COEHA model. Promising examples include: the longer *interstate* conflicts last, the less of an advantage democracies have (Bennett and Stam 1998), the worse the terms of peace are for initiators (Slantchev 2004) and the less likely countries at territorial peace (including democracies) are to win (Gibler and Miller 2012); the later legislators announce their positions on a controversial agenda such as NAFTA (after receiving side payments), the more likely they are to vote for it (Boehmke 2006); and the later the FDA approves a new drug (but just before deadlines), the higher the rates of postmarket safety problems (Carpenter et al. 2012).

This article is organized as follows. The next section introduces the COEHA model. The third section reanalyzes the data of Cunningham et al. (2009). The final section summarizes the argument and offers extensions of the COEHA model.

# 2. MODEL

## 2.1 Data Generation Process

Suppose an observation of duration (e.g. civil war) ends with an event $Z_E$ at time $Z_T > 0$. These two random variables are latent and continuous. The cumulative distribution functions (CDFs) and, if any, probability density functions (PDFs) of their marginal and joint distributions are denoted by

$$
\begin{aligned}
Z_T &\sim F_T(z_T), & \frac{\partial F_T(z_T)}{\partial z_T} &\equiv f_T(z_T) \\
Z_E &\sim F_E(z_E), & \frac{\partial F_E(z_E)}{\partial z_E} &\equiv f_E(z_E) \\
(Z_T, Z_E) &\sim F_{TE}(z_T, z_E), & \frac{\partial^2 F_{TE}(z_E, z_T)}{\partial z_E \partial z_T} &\equiv f_{TE}(z_E, z_T).
\end{aligned}
$$

The continuous event variable $Z_E$ is coarsened into an ordered discrete event variable $Y_E$ as follows. Let $\kappa(-1) \equiv -\infty$ and $\kappa(y_E)$'s be cut points indexed by a non-negative integer $y_E$, so that $\kappa(y_E) \geq \kappa(y_E - 1)$. If any, $\kappa(\max(Y_E)) \equiv \infty$. $Y_E$ takes the value of $y_E$ such that $\kappa(y_E) \geq Z_E > \kappa(y_E - 1)$. For example, if $Z_E$ stands for how advantaged the rebel is in civil war, outcomes may be government victory ($y_E = 0$), settlement ($y_E = 1$), and rebel victory ($y_E = 2$).

Suppose that the censoring time is $z_T^0$ and $Y_T \equiv \min(Z_T, z_T^0)$. If a realized value of time is earlier than the censoring time ($z_T \leq z_T^0$), duration is not censored and one observes a realized value of the ordered discrete event, $y_E$, at time $y_T = z_T$. Otherwise, duration is censored and one observes time $y_T = z_T^0$ but not event $y_E$, which takes on a missing value. Suppose that $D(y_E)$ is a dummy variable which is equal to 0 if $y_E$ is a missing value (that is, the observation is censored) and 1 otherwise.

## 2.2 Copula

$F_{TE}$ remains to be explained. By construction, the quantile variables of time and event, $U_T$ and $U_E$, follow the standard uniform distribution:

$$\begin{aligned} F_T(Z_T) &\equiv U_T \sim \mathcal{U}(u_T) = u_T \\ F_E(Z_E) &\equiv U_E \sim \mathcal{U}(u_E) = u_E. \end{aligned}$$

The CDF of the joint distribution of time and events can be expressed as copula ($C$), a function of their quantile values:

$$F_{TE}(z_T, z_E) \equiv C(u_T, u_E|\theta_{TE}),$$

where $\theta_{TE}$ is the dependence parameter (or such a vector). According to Sklar's Theorem, $C$ is unique. This paper utilizes six copulas: Gaussian, Farlie-Gumbel-Morgenstern (FGM), Frank, Ali-Mikhail-Haq (AMH), Clayton and Gumbel (for details, see Appendix A, Nelsen (2006), and Trivedi and Zimmer (2007)). Example contour plots of their probability density are displayed in Figure 1 where the relatively higher density, the darker gray. (In all panels of this figure, Kendall's $\tau$, a measure of dependence, remains to be 0.18, which is the same as that of the copula estimated in the Application section below. I use R Development Core Team (2010) for drawing figures, analyzing data and conducting simulation.)

The choice of copulas might have substantive implications. For instance, on the one hand, the literature on civil war almost reaches the consensus that, the relatively shorter the war (smaller $u_T$), the more likely is a government victory (smaller $u_E$), because rebels have less military capability than the government at the initial stages of the conflict. This implies strong dependence at the left tail and corresponds to a dense area in the bottom left

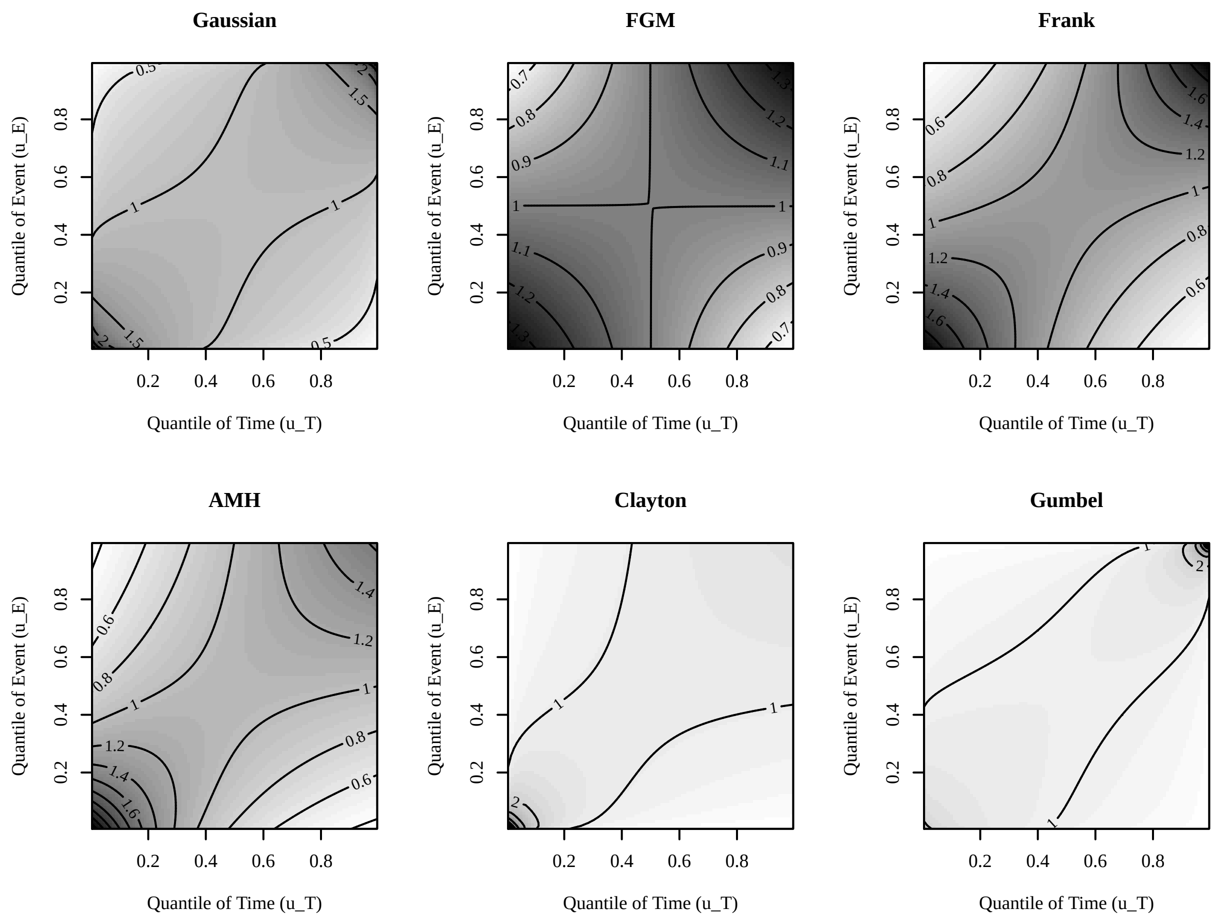


Figure 1: Example contour plots of the probability density of copulas, $\partial^2 C(u_E, u_T)/\partial u_E \partial u_T$. Kendall's $\tau$ is 0.18. The relatively higher density, the darker gray.

corner of the contour plots, except that of the Gumbel copula (Figure 1). On the other hand, scholars do not agree on the case of long internal conflicts. Some studies show that longer wars lead to rebel victory (Thyne 2009, p. 145), because rebels have time to mobilize human and material resources. This means strong dependence at the right tail, namely, a dense area in the top right corner of the contour plots such as the Gaussian (strong dependence), FGM (weak dependence), Frank (middle level dependence) and Gumbel copulas (asymmetric dependence). Other research, however, demonstrates that longer wars result in settlement

(Mason et al. 1999; DeRouen and Sobek 2004) because, for both parties, the accumulated cost outweighs the benefit of would-be victory and settlement is preferred to a continuation of war. In this case, the contour plots of the AMH (a little asymmetric dependence) or Clayton (very asymmetric dependence) copulas seem to be more appropriate, where dependence at the right tail is weak. Sometimes, even within a single article, both lines of argument are mixed (Brandt et al. 2008, 416, 420).

We can also make associated copulas. Suppose that $(U_T, U_E) \sim C(u_T, u_E)$. The CDF of the joint distribution of $\overline{U}_E \equiv 1 - U_E$ and $\overline{U}_T \equiv 1 - U_T$ is called a survival copula;

$$(\overline{U}_T, \overline{U}_E) \sim \overline{C}^{(TE)}(\overline{u}_T, \overline{u}_E) = \overline{u}_T + \overline{u}_E - 1 + C(u_T, u_E).$$

The contour plots of the survival copula is drawn as if we rotate that of the original copula 180 degrees (example contour plots are illustrated in Appendix A). This paper names the copulas of $(U_E, \overline{U}_T)$ and $(\overline{U}_E, U_T)$, time-survival and event-survival copulas, respectively;

$$(\overline{U}_T, U_E) \sim \overline{C}^{(T)}(\overline{u}_T, u_E) \equiv u_E - C(u_T, u_E),$$

$$(U_T, \overline{U}_E) \sim \overline{C}^{(E)}(u_T, \overline{u}_E) \equiv u_T - C(u_T, u_E).$$

Their contour plots are obtained by turning over that of the original copula to the left or upside down. These three copulas are called associated copulas.

## 2.3 Estimation

<u>Parametric Model.</u> $F_T$ is assumed to be a parametric model such as the Weibull distribution. If duration is not censored and both $y_T = z_T$ and $y_E$ are observed, the probability that event

$y_E$ happens at time $y_T$ is

$$
\begin{aligned}
\Pr(y_T = z_T, y_E, D(y_E) = 1) &= \int_{\kappa(y_E-1)}^{\kappa(y_E)} f_{TE}(y_T, z_E) dz_E \\
&= f_T(y_T) \Delta F_{E|T}(\kappa(y_E)|y_T),
\end{aligned}
$$

where

$$
\begin{aligned}
\Delta F_{E|T}(\kappa(y_E)|y_T) &\equiv F_{E|T}(\kappa(y_E)|y_T) - F_{E|T}(\kappa(y_E-1)|y_T) \\
F_{E|T}(\kappa(y_E)|y_T) &\equiv \int_{-\infty}^{\kappa(y_E)} f_{E|T}(z_E|y_T) dz_E \\
&\equiv \int_{-\infty}^{\kappa(y_E)} \frac{f_{TE}(z_E, y_T)}{f_T(y_T)} dz_E \\
&= \frac{\partial C(F_E(\kappa(y_E)), u_T)}{\partial u_T} \\
&\equiv c_{E|T}(u_E(y_E)|u_T). \\
u_E(y_E) &\equiv F_E(\kappa(y_E))
\end{aligned}
$$

I emphasize again that one of the merits of using copulas is their modularity. If we do not employ a copula, it is sometimes difficult or impossible to derive $f_{TE}(z_E, y_T)$ and $f_{E|T}(z_E|y_T)$, and much more so to integrate them. This problem becomes more severe if $F_T(y_T)$ and $F_E(z_E)$ do not belong to the same distribution family, such as exponential, or are not conjugate with each other. By contrast, it is easy and fast to derive, calculate and implement $c_{E|T}(u_E(y_E)|u_T)$ (to be concrete, see Appendix A). If duration is censored and one observes $y_T = z_T^0$ but not $y_E$, the probability that any event $z_E$ would happen at any time $z_T$ after the censoring time $z_T^0$ is the same as the conventional survival probability:

$$
\begin{aligned}
Pr(y_T = z_T^0 < z_T, D(y_E) = 0) &= \int_{y_T}^{\infty} \int_{-\infty}^{\infty} f_{TE}(z_T, z_E) dz_E dz_T \\
&= 1 - F_T(y_T) \\
&\equiv \overline{F}_T(y_T).
\end{aligned}
$$

To sum up, the total likelihood $\mathcal{L}$ of the parameter vector $\boldsymbol{\theta}$ (which is composed of parameters of $F_T$ and $F_E$, $\kappa$'s, and $\theta_{TE}$), given the observed data $y = (y_T, y_E)$, is

$$
\mathcal{L}(\boldsymbol{\theta}|y) \propto \prod_{i=1}^{N} \left[ f_T(y_{Ti}) \times \Delta c_{E|T}(u_{Ei}|u_{Ti}) \right]^{D(y_{Ei})} \left[ \overline{F}_T(y_{Ti}) \right]^{1-D(y_{Ei})},
$$

where $\Delta c_{E|T}(u_E(y_{Ei})|u_{Ti}) = \Delta F_{E|T}(\kappa(y_{Ei})|y_{Ti})$, subscript $i$ corresponds to subject $i$, and $N$ is the number of subjects. (Appendix A shows that the seemingly unrelated discrete-choice duration models (Boehmke 2006) are special cases of the COEHA model.)

Time Varying Covariates. If the values of covariates change during the process, we suppose that an observation is censored and a new observation starts. Suppose that subject (e.g. dyad between government and rebel) $i$'s duration is divided into $n(i)$ observations (e.g. dyad-year) and subject $i$'s observation $j$ (or observation $ij$) lasts from time $y_{Ti(j-1)}$ (when it is left truncated) to time $y_{Tij}$ (when it is censored or, in the case of $j = n(i)$, may have event $y_{Eij}$), where $j = 1, 2, \ldots, n(i)$ and $y_{Ti0} = 0$.

The probability that observation $ij$ is censored at $y_{Tij}$ provided no event has happened by $y_{Ti(j-1)}$ is

$$\begin{aligned}
&Pr(y_{Tij} = z^0_{Tij} < z_{Tij}, D(y_{Eij}) = 0 | y_{Ti(j-1)} < z_{Tij}, \boldsymbol{x}_{Tij}) \\
&= \frac{\overline{F}_T(y_{Tij}|\boldsymbol{x}_{Tij})}{\overline{F}_T(y_{Ti(j-1)}|\boldsymbol{x}_{Tij})} \\
&\equiv \overline{F}_{T|T-1}(y_{Tij}|y_{Ti(j-1)}, \boldsymbol{x}_{Tij}).
\end{aligned}$$

where $\boldsymbol{x}_{Tij}$ is the vector of time varying covariates for $F_T$. The probability that event $y_{Eij}$ happens at $y_{Tij}$, provided no event has happened by $y_{Ti(j-1)}$, is

$$Pr(y_{Tij} = z_{Tij}, y_{Eij}, D(y_{Eij}) = 1 | y_{Ti(j-1)} < z_{Tij}, \boldsymbol{x}_{Tij})$$

$$= f_{T|T-1}(y_{Tij}|y_{Ti(j-1)}, \boldsymbol{x}_{Tij}) \times \Delta c_{E|T}(u_E(y_{Eij})|u_{Tij}),$$

where

$$f_{T|T-1}(y_{Tij}|y_{Ti(j-1)}, \boldsymbol{x}_{Tij}) \equiv \frac{f_T(y_{Tij}|\boldsymbol{x}_{Tij})}{\overline{F}_T(y_{Ti(j-1)}|\boldsymbol{x}_{Tij})}$$

$$u_{Tij} = 1 - \prod_{J=1}^{j} \overline{F}_{T|T-1}(y_{TiJ}|y_{Ti(J-1)}, \boldsymbol{x}_{TiJ})$$

Thus, the total likelihood is

$$\mathcal{L}(\boldsymbol{\theta}|y) \propto \prod_{i=1}^{N}\prod_{j=1}^{n(i)} \Big[f_{T|T-1}(y_{Tij}|y_{Ti(j-1)}, \boldsymbol{x}_{Tij}) \times \Delta c_{E|T}(u_E(y_{Eij})|u_{Tij})\Big]^{D(y_{Eij})}$$
$$\Big[\overline{F}_{T|T-1}(y_{Tij}|y_{Ti(j-1)}, \boldsymbol{x}_{Tij})\Big]^{1-D(y_{Eij})}.$$

Since observations in the same subject are not independent of each other, we employ cluster-pairs bootstrap to estimate standard errors, where a cluster (e.g. conflict involving several parties) is composed of at least one subject (e.g. dyad between parties) and is resampled with replacement.

Cox Model. If one does not assume any parametric model of time, one may use the Cox model. The risk set indicator $R_{ij}(y_T)$ is equal to one if observation $ij$ is at risk of an event at time $y_T$, namely, $y_{Ti(j-1)} < y_T \leq y_{Tij}$, and otherwise zero. Renumber each $i$ in ascending order of $y_{Tin(i)}$. The Cox model assumes the following hazard of observation $ij$ at $y_T$,

$$h_{ij}(y_T) \equiv R_{ij}(y_T) h^0(y_T) \exp(\boldsymbol{x}_{Tij} \boldsymbol{\beta}_T), \tag{1}$$

where $h^0(y_T)$ is the baseline hazard at $y_T$ and $\boldsymbol{\beta}_T$ is the coefficient vector. I propose the COEHA partial likelihood,

$$\mathcal{L}^p(y|\boldsymbol{\theta}) \propto \prod_{i=1}^{N} \left[ f^p_{Tin(i)}(y_{Tin(i)}) \times \Delta c_{E|T}(u_E(y_{Ein(i)})|u_{Tin(i)}) \right]^{D(y_{Ein(i)})}, \tag{2}$$

where the first term is the same as the conventional partial likelihood, that is, the probability that subject $i$ has an event at $y_{Tin(i)}$ conditional on at least one observation having an event at $y_{Tin(i)}$,

$$f^p_{Tin(i)}(y_{Tin(i)}) \equiv \frac{\exp(\boldsymbol{x}_{Tin(i)} \boldsymbol{\beta}_T)}{\sum_{I=1}^{N} \sum_{J=1}^{n(i)} R_{IJ}(y_{Tin(i)}) \exp(\boldsymbol{x}_{TIJ} \boldsymbol{\beta}_T)},$$

and

$$u_{Tij} = 1 - \exp\Big( - \sum_{I=1}^{i} \sum_{J=1}^{j} h_{iJ}(y_{TIn(I)}) D(y_{EIn(I)}) \Big) \tag{3}$$

(Hougaard 2000, 80, Therneau and Grambsch 2000, 266, and Aalen et al. 2008, 141). If a subject has an event at $y_T$, the baseline hazard at $y_T$ is estimated by the Breslow estimate,

$$\hat{h}^0(y_T) \equiv \frac{1}{\sum_{I=1}^{N} \sum_{J=1}^{n(I)} R_{IJ}(y_T) \exp(\boldsymbol{x}_{TIJ} \boldsymbol{\beta}_T)}.$$

By substituting $\hat{h}^0(y_T)$ with $h^0(y_T)$ in Eq. (1) (which is plugged into Eq. (3)) and maximizing Eq. (2), we obtain the estimate of the parameters.

# 3. APPLICATION

Monte Carlo simulation in Appendix B demonstrates that the two currently used methods, separate estimation and CR approaches (explained in the beginning of the article), result in biased estimates if there is dependence among events or dependence between time and events. Moreover, in the case of separate estimation approaches, the bias arises from censoring as well. Based on the results, this paper reanalyzes the data which Cunningham et al. (2009,hereafter, CGS) use in their Table 6. (Their data and STATA code (the original publication version, not the latest updated version) were downloaded on July 6, 2012 from `http://jcr.sagepub.com/content/53/4/570/suppl/DC1`.)

## 3.1 Data

CGS study what influences civil war duration and outcomes using the multinomial logit approach to competing risks (MNL-CR) where censoring is the reference category. For example, they argue that strong rebels tend to lead to shorter wars and decisive outcomes. One of CGS's contributions to the literature is to take into consideration not only state but also non-state attributes such as the rebel's strength. For that purpose, CGS (p. 584) "split conflict involving several parties into different dyads." The data contains 2,201 dyad-year observations from 1946 to 2003 into which CGS divide 327 dyads (which correspond to "subject" in this paper) in 209 conflicts ("cluster" in this paper).Duration is measured in days (not logged). (For four dyads, after one observation ends, about one year lapses before the next observation starts. I calculate the duration time of the 20 observations of these four dyads after the lapse by subtracting the lapse time.)

CGS (p. 587) classify outcomes into four categories, that is, government victory (58 dyads, 18%), rebel victory (40 dyads, 12 %), formal agreement (or negotiated settlement, 65 dyads, 20 %), and low activity ("less than twenty-five deaths per year," 104 dyads, 31 %), while 60 dyads (18 %) are censored. In order to change their unordered discrete event variable into an ordered discrete variable, this paper combines formal agreement and low activity into a new category, "no victory," and rearranges event values in ascending order of government victory ($y_E = 0$), no victory ($y_E = 1$) and rebel victory ($y_E = 2$), where the latent event variable $z_E$ implies the degree of the rebel's advantage (If conditional independence among risks, an assumption of CR models, is correct, then this collapsing procedure should not affect estimates of the covariates' effects on the other events, government and rebel victories. It is the case here as displayed in Appendix C). Below, the COEHA model uses the three category event variable, while my replication of the MNL-CR model employs the four category one as the original analysis does.

In order to focus on what difference the COEHA and MNL-CR models make, the same covariates as CGS (pp. 580-584) are used for time and event regressors in the COEHA model (the median is in parenthesis). (Appendix C elaborates on why we do not have to care about the exclusion restriction.)

- *Territorial Control*: a dummy variable for whether the rebels control territory and exercise a moderate or high level of control over that territory. (0)

- *Rebels Stronger* and *Rebels at Parity*: dummy variables for whether the relative power of rebels to governments is strong (Rebels Stronger) and at parity (Rebels at Parity), respectively, in terms of their ability to fight conventional wars. (0, 0)

- *Legal Political Wing*: a dummy variable for whether a rebel group has an explicit or commonly acknowledged political wing, and whether the political wing is legal. (0)

- *ELF Index*: the national ethnic and linguistic fractionalization index which measures the probability that two randomly selected individuals belong to different ethnic groups. (0.55)

- *Ethnic Conflict*: a dummy variable for whether the conflict pits distinct ethnically-based groups against the state. (0)

- *Ln GDP per Capita*: the natural log of GDP per capita. (7.5)

- *Democracy*: a dummy variable for whether the country has a Polity score equal to or greater than six. (0)

- *Two or More Dyads*: a dummy variable for whether other dyads are active in the same conflict. (1)

- *Ln Population*: the natural log of a country's population. (10.1)

When I replicate CGS's original MNL-CR model, I add to these variables Time at State, "the log of time at state (in days) " (p. 588).

### 3.2 Model Specification

Users of COEHA models must specify marginal distributions of time ($F_T$) and events ($F_E$) as well as the copula function ($C$), though they can do it one by one *separately* thanks to the modularity of COEHA models. This is a great advantage of COEHA models. As for the time model ($F_T$), I employ the Cox model, which assumes proportional hazards but no specific baseline hazard (Eq. (1). CGS also utilize the Cox model in their Tables 3 and 4). Candidates for the events model ($F_E$) are the ordered logit and probit models. When it comes to dependence between time and events ($C$), I consider the six copulas mentioned in the second section and their associated copulas. It turns out that a combination of the ordered logit model and the event-survival AMH copula has the lowest AIC value, 3065.7 (while the AIC values of the Cox model and the ordered logit model are 2590.5 and 481.2, respectively, and their sum is 3071.7, which is larger than the AIC value mentioned above by six. Appendix C displays the AIC values of all of the examined COEHA models). Therefore, I use this model specification below.

### 3.3 Results

Table 1 reports the estimates of the COEHA model.

Dependence between Time and Events. The dependence parameter of the event-survival AMH copula, $\hat{\theta}_{TE}$, is significantly negative. Confusingly, this suggests *positive* dependence between time and events because the *event-survival* AMH copula is used. The left panel of Figure 2 shows the contour plot of the estimated copula. The right panel shows the point

Table 1: Parameter estimates of the COEHA model of Cunningham et al. (2009)'s civil war data (1946-2003)

| | Time ($F_T$) Cox | Event ($F_E$) Ordered Logit |
|---|---|---|
| Territorial Control | $-0.3$ | $-0.3$ |
| | $(-0.7,\ -0.0)$ | $(-1.0,\ 0.5)$ |
| Rebels Stronger | 0.8 | 1 |
| | $(0.3,\ 2.2)$ | $(-1,\ 5)$ |
| Rebels at Parity | 0.4 | 0.9 |
| | $(-0.0,\ 1.0)$ | $(-0.4,\ 2.4)$ |
| Legal Political Wing | 0.5 | $-0.0$ |
| | $(0.0,\ 1.0)$ | $(-0.8,\ 0.8)$ |
| ELF Index | 0.2 | 0.4 |
| | $(-0.4,\ 0.9)$ | $(-1.0,\ 1.9)$ |
| Ethnic Conflict | 0.0 | $-0.2$ |
| | $(-0.3,\ 0.4)$ | $(-0.8,\ 0.5)$ |
| Ln GDP per Capita | 0.05 | $-0.3$ |
| | $(-0.11,\ 0.25)$ | $(-0.6,\ 0.0)$ |
| Democracy | $-0.9$ | 0.7 |
| | $(-1.4,\ -0.4)$ | $(0.1,\ 1.4)$ |
| Two or More Dyads | $-0.4$ | 0.5 |
| | $(-0.7,\ -0.1)$ | $(-0.2,\ 1.3)$ |
| Ln Population | $-0.04$ | $-0.1$ |
| | $(-0.17,\ 0.08)$ | $(-0.4,\ 0.1)$ |
| $\kappa(0)$ | | $-5$ |
| | | $(-8,\ -1)$ |
| $\log(\kappa(1) - \kappa(0))$ | | 1.19 |
| | | $(1.06,\ 1.42)$ |
| | Event-Survival AMH Copula | |
| $\log((1+\theta_{TE})/(1-\theta_{TE}))$ | $-11$ | |
| | $(-13,\ -1)$ | |

NOTE: The Cox model (Eq. (1)) for time ($F_T$), the ordered logit model for event ($F_E$), and the event-survival AMH copula for dependence between time and events ($C$) are employed. The 95% confidence intervals (estimated by cluster-pairs bootstrap) are in parenthesis. The number of dyad-year observations is 2,201. The number of places to the right of the decimal point for a measure is one more than the number of zeros to the right of the decimal point on the standard error of this measure.

estimates of quantiles of observed time ($\hat{u}_T$) for non-censored dyad-year observations on the horizontal axis and that of latent events ($\hat{u}_E$) on the vertical axis. Since the ordered logit model does not estimate $u_E = F_E(z_E)$, but rather $\max(u_E) = F_E(\kappa(y_E))$ and $\min(u_E) = F_E(\kappa(y_E - 1))$, the figure draws segments between the two extrema and plots the midpoints between them. There are few observations in the bottom right corner, which is consistent with positive dependence between time and events.

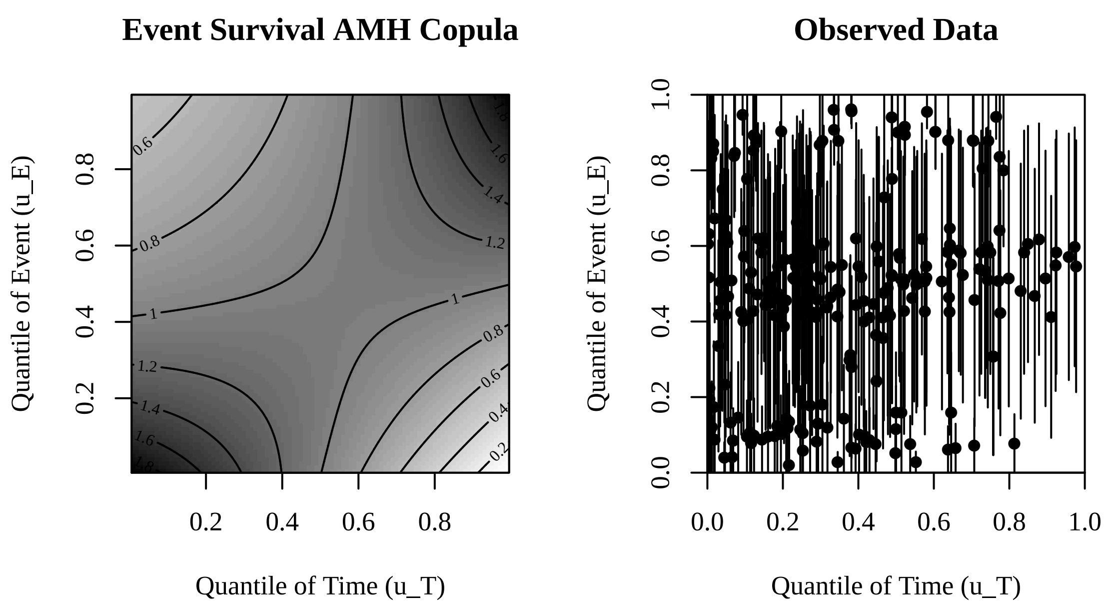


Figure 2: The left panel is the contour plot of the estimated copula. The higher density, the darker gray. The right panel shows the point estimates of quantiles of observed time ($\hat{u}_T$) for 267 non-censored dyad-year observations on the horizontal axis and that of latent events ($\hat{u}_E$) on the vertical axis. Each bar is a segment between the estimated cut-points ($\hat{u}_E(y_E)$ and $\hat{u}_E(y_E - 1)$), and each point is the midpoint between the two cut-points of each segment.

To be concrete, positive dependence between time and events implies that, if civil wars continue longer, rebel victory becomes more likely. Since the parameters of the COEHA model and the MNL-CR model by CGS (their Table 6) are not comparable, Figure 3 shows both models' predictions of the three event probabilities conditional on one type of event happening. I set all covariates at their median values (in the first subsection of this section) as CGS do in their Figure 1, which I call "the baseline scenario," and derive confidence intervals of both models by bootstrap. In this and the following figures, the bold lines and the shaded areas are the point estimates and the 95% confidence intervals in the case of the COEHA model, while the points and the vertical bars are the case of the MNL-CR model. On the one hand, the MNL-CR model suggests that, as civil wars last longer, government victory becomes less likely (left panel), no victory becomes more likely (middle panel, which shows the sum of the probability of formal agreement and that of low activity), and the probability of rebel victory does not change much (right panel). On the other hand, the COEHA model implies that longer civil wars lead to less likely government victory and more likely rebel victory but do not affect the probability of no victory. In addition, the confidence intervals of the COEHA model are smaller than those of the MNL-CR model.

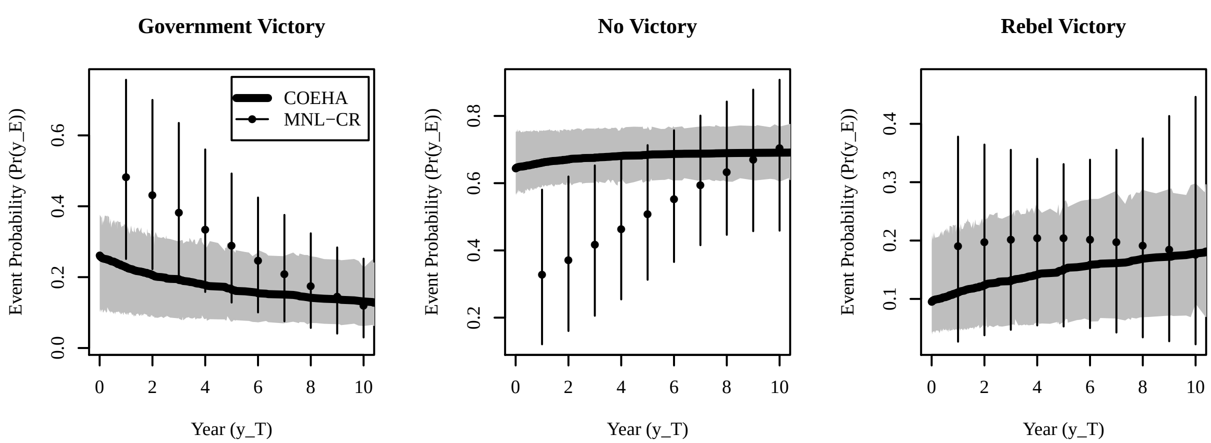


Figure 3: Event probability against Time at State (in years, not logged), conditional on one type of event happening ($Pr(y_E|D(y_E) = 1, y_T)$) in the baseline scenario where all covariates are at their median values.

First Differences. The covariate of the most interest is Rebels Stronger. In the MNL-CR model, the coefficient of Rebels Stronger on rebel victory is significantly positive and its coefficients on all other events are insignificant. According to the COEHA model, however, the coefficient of Rebels Stronger on the latent event variable ($Z_E$, i.e. the advantage of the rebels) is insignificant. Figure 4 shows first differences obtained by subtracting the event probability of the baseline scenario (Rebels Stronger=0) from that of the stronger rebel scenario which is the same as the baseline scenario except Rebels Stronger=1 (the dotted line indicates zero first difference). The MNL-CR first differences in rebel victory probability are significantly positive for several years, while none of the COEHA first differences are. Taking into consideration the results of the Monte Carlo simulation in Appendix B, I suspect that the positive effect of Rebels Stronger on rebel victory in the MNL-CR model is biased upward by dependence between time and events. Substantively speaking, stronger rebels *seem* to be more likely to win, say, in the second year than weak rebels, because wars with stronger rebels should end earlier (that is, the coefficient of Rebels Stronger on the latent time variable ($Z_T$) is positive) and the prolonged conflict (large $u_T$) reveals unexpected weakness of the state (large $u_E$). Note that stronger rebels do not directly raise the marginal probability of rebel victory (that is, the coefficient of Rebels Stronger on the latent event variable ($Z_E$) is zero). In the first place, CGS (p. 577) argue that "wars involving strong rebels will make a decisive military outcome more likely, *either for the rebels or for the government.* Clearly, rebels are more likely to prevail outright the stronger they are, but also ... their forces can be more easily targeted ... with outcomes determined through conventional military tactics or formal negotiations, rather than rooting out clandestine insurgents in the underground" (emphasis in the original). Actually, this "rather paradoxical finding" (DeRouen and Sobek 2004, p. 314) has been repeatedly confirmed in the literature (e.g. Mason et al. 1999; Akcinaroglu 2012). DeRouen and Sobek (2004, p. 314) explain that "a powerful army ... may simply push more people into the arms of the insurgents and send them into hiding." Thus, my null result on the Rebels Stronger effect supports the original authors' hypothesis, though they themselves fail to find evidence for it.

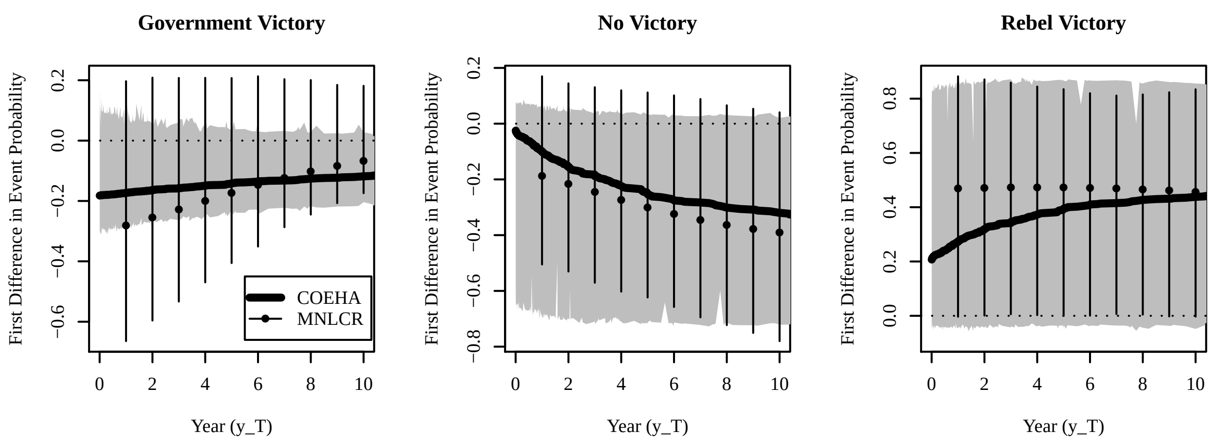


Figure 4: First differences of Rebels Stronger on event probability. The vertical axis shows how much each event probability (conditional on one type of event happening, $Pr(y_E|D(y_E) = 1, y_T)$) changes if Rebels Stronger changes from 0 to 1 with all the other covariates at their median values. The horizontal axis corresponds to Time at State (in years, not logged).

The COEHA model does not always fail to find significant effects when the MNL-CR model does find them. According to the COEHA model, the first differences of Democracy on government and rebel victory probabilities are significantly negative and positive for the first two years (during which almost half of non-censored dyads end), respectively (Figure 5). The MNL-CR model, however, implies that the first differences in government victory probabilities are significantly negative only in the sixth and seventh years and the first differences in rebel victory probabilities are (though insignificantly) even negative after the third year. Moreover, the 95% confidence intervals of the COEHA model are smaller than those of the MNL-CR model.

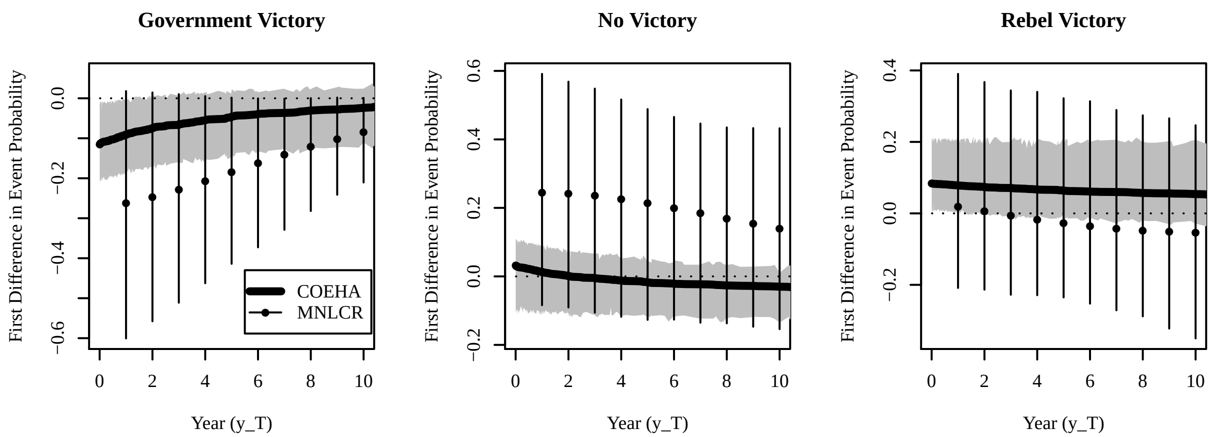


Figure 5: First differences of Democracy on event probability. The vertical axis shows how much each event probability (conditional on one type of event happening, $Pr(y_E|D(y_E) = 1, y_T)$) changes if Democracy changes from 0 to 1 with all the other covariates at their median values. The horizontal axis corresponds to Time at State (in years, not logged).

# 4. CONCLUSION

Scholars are sometimes interested in what event happens and when it happens. If there is dependence among events or dependence between time and events, however, the currently common methods, separate estimation and CR approaches, produce biased estimates. To deal with these problems, this article proposes the COEHA model and demonstrates its usefulness by applying it to a dataset of civil wars. The results contribute to two substantive questions as well. First, as wars become longer, rebels are more likely to win but settlement is no more or less likely; second, stronger rebels lengthen wars but do not necessarily tend to win, as the original authors predict but their own analysis using the CR model fails to establish. It also turns out that, in democracies, civil wars are longer and the probability of government (or rebel) victory is lower (or higher) than in non-democratic countries, especially if intrastate conflicts last for less than a few years.

More importantly, we can extend the COEHA model to a larger class of event history analysis. First, we can deal with sample selection (Boehmke et al. 2006) and informative censoring (Huang and Zhang 2008) by making $y_E$ a dummy variable for whether an observation is selected or censored. Second, if an event variable is continuous, we only have to suppose that, if an observation is not censored, we observe the latent event $z_E$ and substitute the likelihood,

$$\mathcal{L}(\boldsymbol{\theta}|y) \propto p(y_T = z_T, y_E = z_E) = f_{TE}(y_T, y_E) = f_T(y_T) f_E(y_E) \frac{\partial^2 C(u_E, u_T|\theta_{TE})}{\partial u_T \partial u_E}. \quad (4)$$

For instance, the longer the civil war, the more deaths (Moore 2012); the later the election, the more votes parties in the cabinet garner (Smith 2004); the longer a party is out of office, the more impatient it is and the smaller its share of ministerial posts obtained (Falcó-Gimeno 2012); or the longer workers strike, the larger the wage increase (Card and Olson 1995). Finally, the dependent CR models (Gordon 2002; Fukumoto 2009), and the shared frailty models (Box-Steffensmeier and Jones 2004, pp. 162-166) can be derived and extended by incorporating them into Eq. (4) where one regards $Z_E$ as another duration. In the interdependent duration model, if the second duration is censored, the COEHA model holds where $Y_E$ is a dummy variable for whether duration is censored or not.

All told, I hope that the COEHA model, its extensions and any copula-based models are helpful for various empirical applications and methodological issues.

# SUPPLEMENTARY MATERIALS

All of the following supplemental files are contained in the file "supplementary_materials.zip".

**Appendices:** Appendix A elaborates on copulas and shows that the seemingly unrelated discrete-choice duration models (Boehmke 2006) are special cases of the COEHA model. Appendix B demonstrates Monte Carlo simulation results. Appendix C illustrates additional analyses of application. ("appendices.pdf")

**Replication Files:** Following the instruction in the file "readme.txt", readers should be able to replicate all tables and figures in the manuscript and the appendices. ("replication" folder)

# References

Aalen, O. O., Borgan, Ø., and Gjessing, H. K. (2008), *Survival and Event History Analysis: A Process Point of View*, New York, NY: Springer.

Akcinaroglu, S. (2012), "Rebel Interdependencies and Civil War Outcomes," *Journal of Conflict Resolution*, 56, 879–903.

Balch-Lindsay, D., Enterline, A. J., and Joyce, K. A. (2008), "Third-Party Intervention and the Civil War Process," *Journal of Peace Research*, 45, 345–363.

Bennett, S. D. and Stam, A. C. (1998), "The Declining Advantages of Democracy: A Combined Model of War Outcomes and Duration," *Journal of Conflict Resolution*, 42, 344–366.

Boehmke, F. J. (2006), "The Influence of Unobserved Factors on Position: Timing and Content in the NAFTA Vote," *Political Analysis*, 14, 421–438.

Boehmke, F. J., Morey, D. S., and Shannon, M. (2006), "Selection Bias and Continuous-Time Duration Models: Consequences and a Proposed Solution," *American Journal of Political Science*, 50, 192–207.

Box-Steffensmeier, J. M. and Jones, B. S. (2004), *Event History Modeling: A Guide for Social Scientists*, Cambridge, UK: Cambridge University Press.

Brandt, P. T., Mason, T. D., Gurses, M., Petrovsky, N., and Radin, D. (2008), "When and How the Fighting Stops: Explaining the Duration and Outcome of Civil Wars," *Defence and Peace Economics*, 19, 415–434.

Card, D. and Olson, C. A. (1995), "Bargaining Power, Strike Duration, and Wage Outcomes: An Analysis of Strikes in the 1880s," *Journal of Labor Economics*, 13, 32–61.

Carpenter, D., Chattopadhyay, J., Moffitt, S., and Nall, C. (2012), "The Complications of Controlling Agency Time Discretion: FDA Review Deadlines and Postmarket Drug Safety," *American Journal of Political Science*, 56, 98–114.

Cunningham, D. E., Gleditsch, K. S., and Salehyan, I. (2009), "It Takes Two: A Dyadic Analysis of Civil War Duration and Outcome," *Journal of Conflict Resolution*, 53, 570–597.

DeRouen, Karl R., J. and Sobek, D. (2004), "The Dynamics of Civil War Duration and Outcome," *Journal of Peace Research*, 41, 303–320.

Falcó-Gimeno, A. (2012), "Parties Getting Inpatient: Time Out of Office and Portfolio Allocation in Coalition Governments," *British Journal of Political Science*, 42, 393–411.

Fearon, J. D. (2004), "Why Do Some Civil Wars Last so Much Longer than Others?" *Journal of Peace Research*, 41, 275–301.

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

Gibler, D. M. and Miller, S. V. (2012), "Quick Victories?: Territory, Democracies, and Their Disputes," *Journal of Conflict Resolution*, 57, 258–284.

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

Hougaard, P. (2000), *Analysis of Multivariate Survival Data*, New York, NY: Springer.

Huang, X. and Zhang, N. (2008), "Regression Survival Analysis with an Assumed Copula for Dependent Censoring: A Sensitivity Analysis Approach," *Biometrics*, 64, 1090–1099.

Mason, T. D. and Fett, P. J. (1996), "How Civil Wars End: A Rational Choice Approach," *Journal of Conflict Resolution*, 40, 546–568.

Mason, T. D., Weingarten, Jr., J. P., and Fett, P. J. (1999), "Win, Lose, or Draw: Predicting the Outcome of Civil Wars," *Political Research Quarterly*, 52, 239–268.

Moore, M. (2012), "Selling to Both Sides: The Effects of Major Conventional Weapons Transfers on Civil War Severity and Duration," *International Interactions*, 38, 325–347, moore, Matthew.

Nelsen, R. B. (2006), *An Introduction to Copulas*, New York, NY: Springer, 2nd ed.

R Development Core Team (2010), *R: A Language and Environment for Statistical Computing*, Vienna: R Foundation for Statistical Computing.

Regan, P. M. (2002), "Third-Party Interventions and the Duration of Intrastate Conflicts," *Journal of Conflict Resolution*, 46, 55–73.

Slantchev, B. L. (2004), "How Initiators End Their Wars: The Duration of Warfare and the Terms of Peace," *American Journal of Political Science*, 48, 813–829.

Smith, A. (2004), *Election Timing*, Cambridge, UK: Cambridge University Press.

Therneau, T. M. and Grambsch, P. M. (2000), *Modeling Survival Data: Extending the Cox Model*, New York, NY: Springer.

Thyne, C. L. (2009), *How International Relations Affect Civil Conflict: Cheap Signals, Costly Consequences*, Lanham, MD: Lexington Books.

Trivedi, P. K. and Zimmer, D. M. (2007), *Copula Modeling: An Introduction for Practioners*, Boston, MA: Publishers Inc.