Appendices of

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

Kentaro Fukumoto †

---

* *Journal of American Statistical Association* forthcomig.

†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.

APPENDIX A: DETAILS OF THE MODELS

A.1 Copula

The most straightforward copula is Gaussian:

$$C^{(G)}(u_T, u_E|\theta_{TE}) \equiv \Phi_2(\Phi^{-1}(u_T), \Phi^{-1}(u_E)|\theta_{TE}),$$

where Φ and Φ_2 are the CDFs of the univariate and bivariate standard normal distributions, respectively, and $-1 < \theta_{TE} < 1$ represents the correlation coefficient.

The next is the Farlie-Gumbel-Morgenstern (FGM) copula:

$$C^{(FGM)}(u_T, u_E|\theta_{TE}) \equiv u_T u_E(1 + \theta_{TE}\overline{u}_T\overline{u}_E),$$

where $-1 < \theta_{TE} < 1$. Compared to the Gaussian copula, since dependence is weaker (or the slope is less steep) at both tails, the dark gray (or relatively higher density) area is larger. (For example, the top right corner density of the FGM copula is relatively higher than the bottom right corner density of the FGM copula, but not absolutely higher than the top right corner density of the Gaussian copula.)

Archimedean Copula. Archimedean copulas are a class of copulas for which there is a generator function $g(u)$ such that

$$C^{(A)}(u_T, u_E|g(u)) \equiv g^{-1}(g(u_T) + g(u_E)).$$

An example is the Frank copula. Its generator function is

$$g^{(F)}(u|\theta_{TE}) \equiv -\log\frac{\exp(-u\theta_{TE}) - 1}{\exp(-\theta_{TE}) - 1}.$$

Dependence is weaker at both tails than in the Gaussian copula but stronger than in the FGM copula.

The second Archimedean copula is the Ali-Mikhail-Haq (AMH) copula whose generator function is

$$g^{(AMH)}(u|\theta_{TE}) \equiv \log \frac{1-\bar{u}\theta_{TE}}{u},$$

where $-1 < \theta_{TE} < 1$ and $\bar{u} = 1 - u$. Dependence is strong at the left tail but weak at the right tail in this example, though the dependence structure varies with θ_{TE}.

Third, the Clayton copula is derived by way of Archimedean representation with the generator function

$$g^{(C)}(u|\theta_{TE}) \equiv \frac{u^{-\theta_{TE}}-1}{\theta_{TE}},$$

where $\theta_{TE} > 0$. This copula is appropriate if dependence is very strong at the left tail but very weak at the right tail. It cannot capture negative dependence.

Finally, the Gumbel copula's generator function is

$$g^{(G)}(u|\theta_{TE}) \equiv (-\log(u))^{\theta_{TE}},$$

where $\theta_{TE} > 1$. Contrary to the AMH and Clayton copulas, dependence is weak at the left tail but strong at the right tail.

Associated Copula. Figure A.1 illustrates the contour plots of the associated copulas of the Clayton copula.

First Derivative of Copula. In the case of an Archimedean copula,

$$c^{(A)}_{E|T}(u_E(y_E)|u_T, g(u)) = \frac{g'(u_T)}{g'(C^{(A)}(u_T, u_E(y_E)|g(u)))}$$

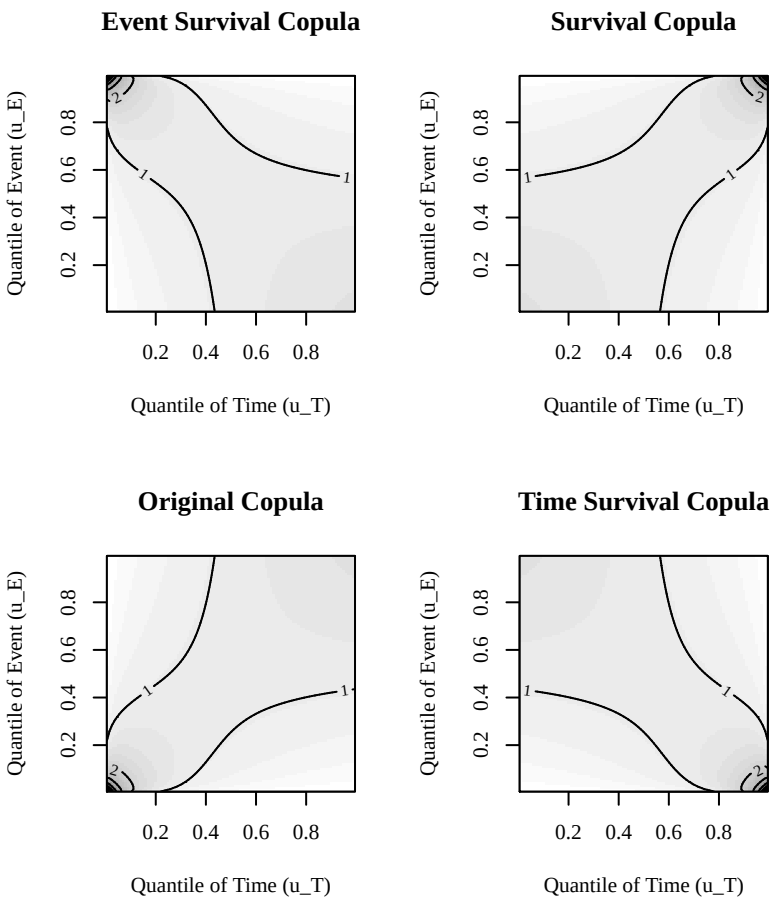

Figure A.1: Example contour plots of the probability density of associated copulas of the Clayton copula.

(Trivedi and Zimmer 2007, p. 43). Thus, in the case of the AMH copula (see also Ali et al. 1978, p. 409),

$$c_{E|T}^{(AMH)}(u_E(y_E)|u_T, g^{(AMH)}(u|\theta_{TE})) = \frac{u_E(y_E)(1-\theta_{TE}\overline{u}_E(y_E))}{(1-(\theta_{TE}\overline{u}_T\ \overline{u}_E(y_E))^2}.$$

where $\overline{u}_E(y_E) = 1 - u_E(y_E)$.

A.2 Seemingly Unrelated Discrete-Choice Duration Models

To my knowledge, the only other models to take into account dependence between time and events are the two seemingly unrelated discrete-choice duration (SUDCD) models (Boehmke 2006). Nevertheless, these are special cases of the COEHA model.

The log-normal SUDCD model denotes

$$\begin{aligned} z_T^{*(LN)} &\equiv \tfrac{\log(z_T)-\boldsymbol{x}_T\boldsymbol{\beta}_T}{\alpha_T} \\ z_E^{*(LN)} &\equiv z_E - \boldsymbol{x}_E\boldsymbol{\beta}_E, \end{aligned}$$

where $\boldsymbol{x}_T$ and $\boldsymbol{x}_E$ are covariate vectors for time and events, respectively, $\boldsymbol{\beta}_T$ and $\boldsymbol{\beta}_E$ are their coefficient vectors, and $\alpha_T > 0$ is an ancillary parameter. The model assumes that these transformed latent time and event variables follow the standard bivariate normal distribution,

$$F_{TE}^{(SUDCD-LN)}(z_T, z_E) \equiv \Phi_2\Big(z_T^{*(LN)}, z_E^{*(LN)}\Big|\theta_{TE}\Big),$$

and, if $z_E^{*(LN)} < \kappa(0)$, $y_E = 0$; otherwise $y_E = 1$. It is straightforward to rewrite this in the COEHA form in which F_T is the log normal distribution, F_E is the normal distribution, and C is the Gaussian copula.

$$F_{TE}^{(SUDCD-LN)}(z_T, z_E) = C^{(G)}(\Phi(z_T^{*(LN)}), \Phi(z_E^{*(LN)})|\theta_{TE}).$$

Next, the Weibull SUDCD model denotes

$$\begin{aligned} z_T^{*(W)} &\equiv \exp\left(\tfrac{\log(z_T)-\boldsymbol{x}_T\boldsymbol{\beta}_T}{\alpha_T^{-1}}\right) \\ z_E^{*(W)} &\equiv \exp(\log(z_E) - \boldsymbol{x}_E\boldsymbol{\beta}_E), \end{aligned}$$

and assumes that these transformed latent time and event variables follow the bivariate exponential distribution,

$$F_{TE}^{(SUDCD-W)}(z_T, z_E) \equiv (1-\exp(-z_T^{*(W)}))(1-\exp(-z_E^{*(W)}))(1+\theta_{TE}\exp(-z_T^{*(W)} - z_E^{*(W)})),$$

and, if $z_E^{*(W)} < \kappa(0)$, $y_E = 0$; otherwise $y_E = 1$. Here z_T follows the Weibull distribution, while z_E follows the exponential distribution. Therefore, the Weibull SUDCD model is represented by the COEHA model where F_T is the Weibull distribution, F_E is the exponential distribution, and C is the FGM copula.

$$F_{TE}^{(SUDCD-W)}(z_T, z_E) = C^{(FGM)}(F_T^{(W)}(z_T), F_E^{(W)}(z_E)|\theta_{TE}).$$

The SUDCD models specify marginal distributions of events and time as the marginal distributions of the known bivariate distribution (such as normal, exponential and their same family distributions) so that analysts can easily derive the joint distribution of time and events. This consideration of computation, however, restricts the choice of marginal distributions, while the real data generation process does not have to care about mathematical convenience. By contrast, the COEHA model can combine *any* kind of marginal distributions with *any* type of copula to obtain *any* kind of joint distribution. This flexibility or modularity is a great advantage of the COEHA model.

A.3 Extensions of the COEHA Model

The conclusion mentions a larger class of models which "can be derived and extended ... where one regards Z_E as another duration." This subsection elaborates on them.

Interdependent Duration Model. Suppose that subject i has two latent durations (or spells) whose lengths are denoted by random variables Z_1 and Z_2. Assume

$$\begin{aligned} Z_k &\sim F_k(z_k) && \equiv u_k \quad (k = 1, 2) \\ (Z_1, Z_2) &\sim F_{12}(z_1, z_2) && \equiv C(u_1, u_2). \end{aligned}$$

If $C(u_1, u_2) \neq u_1 u_2$, the two durations are dependent on each other and should be modeled simultaneously as an "interdependent duration model. " If neither duration is censored, the likelihood is

$$\mathcal{L}(\boldsymbol{\theta}|y) \propto p(y_1 = z_1, y_2 = z_2, D(y_1) = D(y_2) = 1) = f_1(y_1) f_2(y_2) \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2}.$$

If the k-th duration ($k = 1, 2$) is censored at y_k but the $\overline{k}$-th duration ($\overline{k} \equiv 3 - k$) is not, the likelihood is

$$\mathcal{L}(\boldsymbol{\theta}|y) \propto p(y_{\overline{k}} = z_{\overline{k}}, y_k, D(y_{\overline{k}}) = 1, D(y_k) = 0) = f_{\overline{k}}(y_{\overline{k}}) \frac{\partial \overline{C}^{(k)}(u_1, u_2)}{\partial u_{\overline{k}}}.$$

If both durations are censored, the likelihood is

$$\mathcal{L}(\boldsymbol{\theta}|y) \propto \Pr(y_1, y_2, D(y_1) = D(y_2) = 0) = \overline{C}^{(12)}(u_1, u_2)$$

Dependent Competing Risks Model. I continue to use the notation above. In a (dependent) competing risks model, we observe either z_1 or z_2 at most, but not both. If we observe $y_{\overline{k}} = z_{\overline{k}}$ but the k-th duration is censored at y_k, we think that an event (e.g. the end of a legislative career) is caused by the $\overline{k}$-th risk (e.g. electoral defeat) and not by the k-th risk (e.g. retirement). If both durations are censored, we do not observe any event (e.g. the legislative career continues). The likelihood is the same as that of an interdependent duration model.

Shared Frailty Model. To make a clear connection between a shared frailty model and the use of a copula, I begin with a basic model. Assume that both durations follow log normal distributions given duration ki specific parameters ($\theta_{ki} = (\mu_{ki}, \sigma_k), k = 1, 2$) as well

as subject i's specific shared frailty (ν_i) and the frailty follows a normal distribution given parameter $\theta_{12} = \sigma_{12}$;

$$\begin{aligned} Z_{ki}|\theta_{ki}, \nu_i &\sim F_k^{(LN)}(z_{ki}|\theta_{ki}, \nu_i) \equiv \Phi\Big(\tfrac{\log(z_{ki})-(\mu_{ki}+\nu_i)}{\sigma_k}\Big) \quad (k = 1, 2) \\ \nu_i|\theta_{12} &\sim F_\nu^{(N)}(\nu_i|\theta_{12}) \equiv \Phi\Big(\tfrac{\nu_i}{\sigma_{12}}\Big) \end{aligned}$$

It follows that, given θ_{1i}, θ_{2i}, and θ_{12}, but not the shared frailty ν_i itself, the joint distribution of Z_{1i} and Z_{2i} follows a bivariate log normal distribution;

$$\begin{aligned} (Z_{1i}, Z_{2i})|\theta_{1i}, \theta_{2i}, \theta_{12} &\sim F_{12}^{(SharedFrailty-LN)}(z_{1i}, z_{2i}|\theta_{1i}, \theta_{2i}, \theta_{12}) \\ &= \int F_1^{(LN)}(z_{1i}|\theta_{1i}, \nu_i) F_2^{(LN)}(z_{2i}|\theta_{2i}, \nu_i) \Phi\Big(\frac{\nu_i}{\sigma_{12}}\Big) d\nu_i \\ &= \Phi_2(z_{1i}^{*(LN)}, z_{2i}^{*(LN)}|\theta_{12}^*) \end{aligned}$$

where

$$\begin{aligned} z_{ki}^{*(LN)} &= \frac{\log(z_{ki}) - \mu_{ki}}{\sqrt{\sigma_k^2 + \sigma_{12}^2}} \quad (k = 1, 2) \\ \theta_{12}^* &= \frac{\sigma_{12}^2}{\sqrt{\sigma_1^2 + \sigma_{12}^2}\sqrt{\sigma_2^2 + \sigma_{12}^2}}. \end{aligned}$$

It turns out that this joint distribution can be expressed by using the Gaussian copula;

$$F_{12}^{(SharedFrailty-LN)}(z_{1i}, z_{2i}|\theta_{1i}, \theta_{2i}, \theta_{12}) = C^{(G)}(\Phi(z_{1i}^{*(LN)}), \Phi(z_{2i}^{*(LN)})|\theta_{12}^*).$$

In similar ways, we can express any other shared frailty models by using marginal distributions $F_k(z_{ki})$ other than the log normal and another copula $C(u_{1i}, u_{2i})$ different from the Gaussian.

APPENDIX B: MONTE CARLO SIMULATION

If dependence among events or dependence between time and events are present, models which fail to take them into account produce biased estimates. Moreover, in the case of separate estimation approaches, the bias arises from censoring as well. The present appendix demonstrates these with Monte Carlo simulation.

B.1 Setup

Fake datasets are generated by the following COEHA model. The setup is constructed so that the results are helpful for the Application section in the manuscript. The number of subjects, N, is 300. Covariates x_1 and x_2 are randomly drawn from the standard normal distribution once, and then used in all of the simulations. Let $\boldsymbol{x}_T = (1, x_1)$ and $\boldsymbol{x}_E = (x_1, x_2)$.

For subject i, I generate u_{Ti} randomly from the standard uniform distribution, $\mathcal{U}$, and make $z_{Ti} = F_T^{(W)-1}(u_{Ti}|\boldsymbol{x}_{Ti}, \alpha_T, \boldsymbol{\beta}_T)$ where $F_T^{(W)}(z_T|\boldsymbol{x}_T, \alpha_T, \boldsymbol{\beta}_T) \equiv 1-\exp(-(z_T \exp(\boldsymbol{x}_T\boldsymbol{\beta}_T))^{\alpha_T})$ is the Weibull distribution, $\alpha_T = -1$ and $\boldsymbol{\beta}_T = (\beta_{T0}, \beta_{T1}) = (-1, 5)$. If $z_{Ti} > z_T^0 = 50$, the observation is censored and let $y_{Ti} = z_T^0$ and y_{Ei} is a missing value. Otherwise, I make $y_{Ti} = z_{Ti}$, sample $u_{E|Ti}$ randomly from $\mathcal{U}$, calculate $u_{Ei} = c_{E|T}^{(ES-AMH)-1}(u_{E|Ti}|u_{Ti})$ and $z_{Ei} = F_E^{(L)-1}(u_{Ei}|\boldsymbol{x}_{Ei}, \boldsymbol{\beta}_E)$ where $C^{(ES-AMH)}$ is the event-survival AMH copula and $F_E^{(L)}(z_E|\boldsymbol{x}_E, \boldsymbol{\beta}_E) \equiv 1/(1+\exp(-(z_E-\boldsymbol{x}_E\boldsymbol{\beta}_E)))$ is the logistic distribution, $\boldsymbol{\beta}_E = (\beta_{E1}, \beta_{E2}) = (0, 1)$ (the value of θ_{TE} will be explained below). To be concrete,

$$
\begin{aligned}
F_T^{(W)-1}(u_T) &= \frac{(-\log \overline{u}_T)^{\alpha_T^{-1}}}{\exp(\boldsymbol{x}_T\boldsymbol{\beta}_T)} \\
F_E^{(L)-1}(u_E) &= \boldsymbol{x}_E\boldsymbol{\beta}_E + \log \frac{u_E}{\overline{u}_E} \\
c_{E|T}^{(ES-AMH)-1}(u_{E|T}|u_T) &= 1 - \frac{2\overline{u}_{E|T}(\overline{u}_T\theta_{TE}-1)^2}{w_1+\sqrt{w_2}}
\end{aligned}
$$

where

$$w_1 = -\theta_{TE}(2\overline{u}_T\ \overline{u}_{E|T} + 1) + 2\theta_{TE}^2\overline{u}_T^2\overline{u}_{E|T} + 1$$

$$w_2 = \theta_{TE}^2(4\overline{u}_T^2\overline{u}_{E|T} - 4\overline{u}_T\ \overline{u}_{E|T} + 1) + \theta_{TE}(4\overline{u}_{E|T} - 4\overline{u}_T\ \overline{u}_{E|T} - 2) + 1.$$

I convert z_{Ei} to one of three ordered events (y_{Ei}), that is, 0 (Low), 1 (Middle) and 2 (High), by comparing z_{Ei} with $\kappa(0) = -1$ and $\kappa(1) = 0.5$. I repeat the procedure of this paragraph for N subjects.

In order to construct subject-time observations, I repeated subject i's covariates $n(i) = \lceil z_{Ti} \rceil$ times (which is the ceiling of z_{Ti}) so that $\boldsymbol{x}_{Tij} = \boldsymbol{x}_{Ti}, \boldsymbol{x}_{Eij} = \boldsymbol{x}_{Ei}$ (that is, covariates are time invariant). Up to observation $j \leq n(i) - 1$ of subject i, observation ij's duration starts at $y_{Ti(j-1)} = j - 1$ and is censored at $y_{Tij} = j$ and y_{Eij} is a missing value. In the case of the observation $j = n(i)$ of subject i, if the subject is censored, $y_{Tij} = z_T^0 < z_{Ti}$ and y_{Eij} is a missing value. Otherwise, $y_{Tij} = z_{Ti} \leq z_T^0$ and $y_{Eij} = y_{Ei}$ is observed.

Using this fake dataset, I estimate the parameters of five models. The first model is a COEHA model where F_T is the Weibull distribution, F_E is the logistic distribution, C is the event-survival AMH copula, covariates for time and events are $\boldsymbol{x}_T$ and $\boldsymbol{x}_E$, respectively. The next two are separate estimation approaches. They use ordinary event history analysis (with the Weibull distribution) to explain time y_T by $\boldsymbol{x}_T$ and the ordered logit model to explain events y_E in two ways: first, by $\boldsymbol{x}_E$ and $\log(y_T)$ (the second model, Weibull and ordered logit *with* time (OLT)) and, second, by $\boldsymbol{x}_E$ but not $\log(y_T)$ (the third model, Weibull and ordered logit *without* time (OLNOT)). The last two models are CR models (Box-Steffensmeier and Jones 2004, pp. 168-175). The fourth model is the latent survivor time approach to competing risks (LST-CR) where each of three event history analyses regard one type of event as "the"

event but the other types of events as censoring, uses $\boldsymbol{x}_E$ as covariates and assumes the Weibull distribution. The fifth model is the multinomial logit approach to competing risks (MNL-CR) where censoring is the reference category and regressors are $\boldsymbol{x}_E$ and $\log(y_T)$.

I examine 19 cases where θ_{TE} varies from -0.9 to 0.9 by 0.1. For each value of θ_{TE}, I make 1,000 fake datasets of y_T and y_E (not $\boldsymbol{x}_T$ and $\boldsymbol{x}_E$, which are constant across simulations) and calculate the averages and 95% confidence intervals of parameter estimates as well as other quantities of interest (such as outcome probability and first differences). One subject has 14.3 observations on average, though 54 % of subjects have only one observation. The shares of Low, Middle, and High events and censoring are 25 %, 25 %, 28 %, and 22 %, respectively.

B.2 Results

Separate Estimation Approaches. Since the covariates' coefficients of the COEHA model and the two separate estimation approaches denote the same estimands, we only have to compare their estimates directly. Figure B.1 displays $\hat{\beta}_{E1}$ (vertical axis, left panel) and $\hat{\beta}_{E2}$ (right panel) for various values of θ_{TE} (horizontal axis). The bold lines and the shaded areas are the mean and 95% confidence interval of the COEHA estimates; the lines with circles and the shaded areas are the case of the OLT model; and the lines with squares and the two dashed bold lines are the average as well as the upper and lower bounds of the 95% confidence interval of the OLNOT estimates. As a matter of course, the COEHA estimates are unbiased (recall that $\beta_{E1} = 0, \beta_{E2} = 1$).

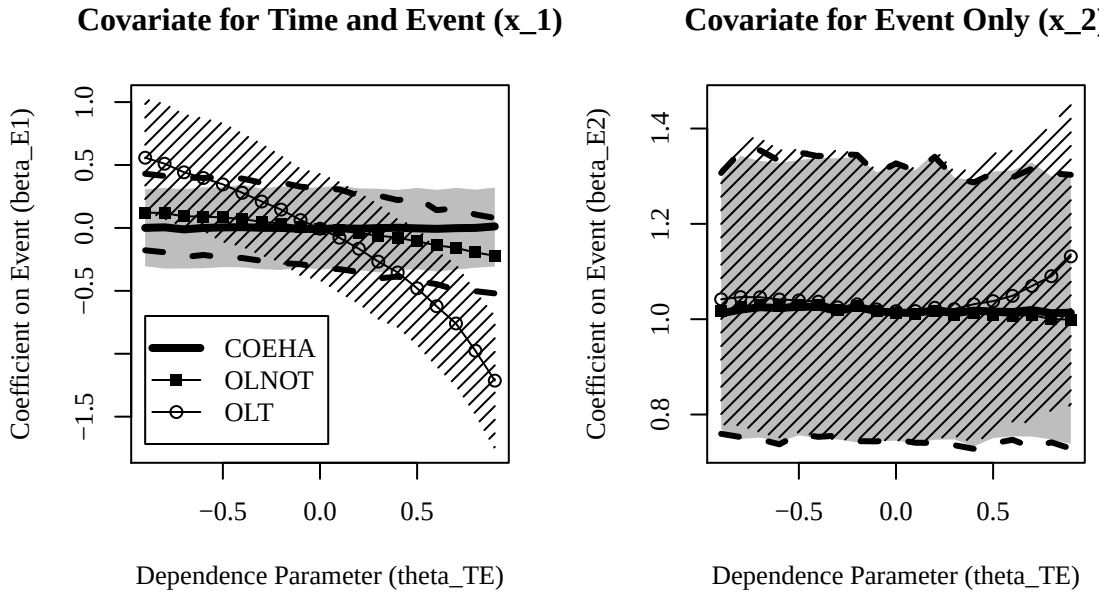

Figure B.1: Coefficients on event estimated by three models against dependence parameter θ_{TE} in the case of regular censoring time ($z_T^0 = 50$). The left and right panels report $\hat{\beta}_{E1}$ and $\hat{\beta}_{E2}$, respectively.

The OLNOT estimates of β_{E1} are slightly biased. This is caused by *censoring*. In the case of $\theta_{TE} > 0$, the larger z_T is, the smaller z_E is but the more likely the observation is to be censored at $z_T^0 = 50$. It follows that observations with large (or small) values of y_E are more (or less) likely to be observed than they are to be generated. This problem is less severe if $x_1 = 1$ (the hazard is higher, z_T is shorter and an observation is less likely to be censored) than if $x_1 = 0$ ($\because \beta_{T1} > 0$). Thus, the larger x_1, the smaller y_E seems to be, leading to a downward bias of $\hat{\beta}_{E1}$. Figure B.2 presents another simulation with a shortened censoring time of $z_T^0 = 2$ (from $z_T^0 = 50$). When it comes to OLNOT estimates of β_{E1} (the left panel) in the case of $\theta_{TE} = 0.9$, their average is -0.41 and their 95% confidence interval ranges from -0.86 to -0.03, which does not cover zero. Thus, the more censored observations, the more

biased are the estimates. If no observations are censored, the OLNOT model represents the true marginal distribution of events. Usually, however, analysts are interested in the relationship between time and events, which the OLNOT model does not deal with.

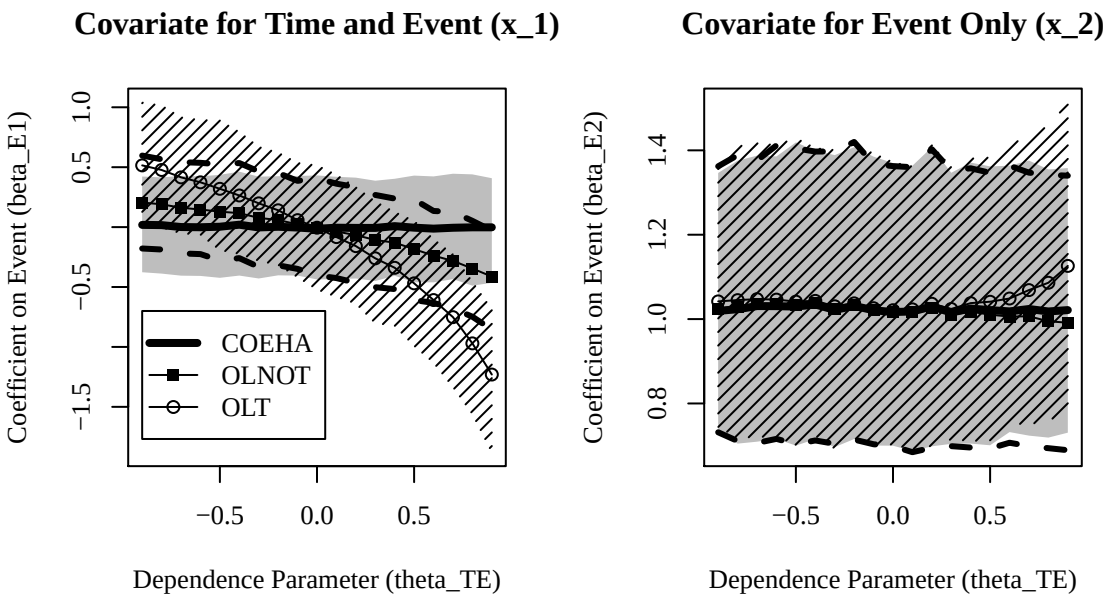

Figure B.2: Coefficients on event estimated by three models against dependence parameter θ_{TE} in the case of shortened censoring time ($z_T^0 = 2$). The left and right panels report $\hat{\beta}_{E1}$ and $\hat{\beta}_{E2}$, respectively.

The OLT estimates of β_{E1} are more biased than the OLNOT estimates unless $\theta_{TE} = 0$. In particular, if the size of θ_{TE} is large, the 95% confidence interval does not cover the true value, zero. The bias arises from *censoring* (as with the OLNOT model) as well as *dependence between time and events*. This is a kind of *endogeneity bias* because $\log(y_T)$ is regarded as exogenous, while it is in fact endogenous. As a result, roughly speaking, the size of $\hat{\beta}_{E1}$'s bias is related to the covariance between x_1 and z_T (or β_{T1}) times the covariance between z_T and z_E (dependence between time and events).

Since both ordered logit models correctly take into account dependence among events, their estimates of β_{E2} are unbiased, even though there is dependence between time and events (except for the OLT estimates in the case of a very large θ_{TE}). According to Figure B.3 (hereafter, $z_T^0 = 50$, again), the estimates of α_T, β_{T0}, and β_{T1} by ordinary event history analysis (with Weibull distribution) are unbiased (crosses) and their 95% confidence intervals (between the two dashed bold lines) almost overlap those of the COEHA model (whose esimates' mean and 95% confidence interval are indicated by the thin lines and the shaded areas). The reason is that ordinary event history analysis correctly assumes the true marginal distribution of z_T, F_T.

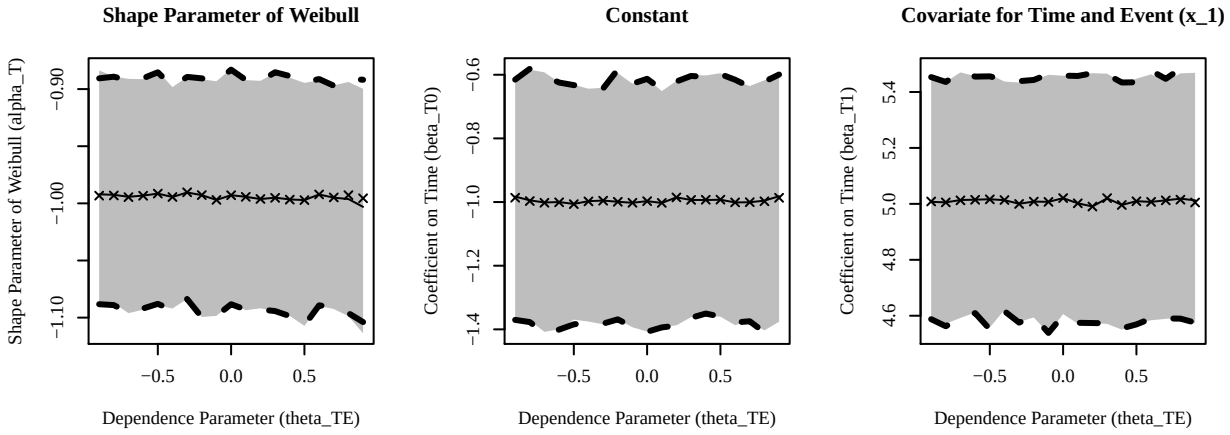

Figure B.3: Coefficients on time ($\hat{\alpha}_T$ (left), $\hat{\beta}_{T0}$ (middle) and $\hat{\beta}_{T1}$ (right)) against dependence parameter θ_{TE} in the case of regular censoring time ($z_T^0 = 50$). The solid lines and the shaded areas are the mean and 95% confidence interval of COEHA estimates. The crosses and dashed lines are the case of ordinary event history analysis (with Weibull distribution).

Competing Risks Approaches. Unlike the separate estimation approaches above, the covariates' coefficients ($\boldsymbol{\beta}_E$ for every event) of the two CR models are not comparable with those (one $\boldsymbol{\beta}_E$) of the COEHA model, because they do not refer to the same estimands.

Thus, we compare not coefficients estimates but quantities of interest such as events' probabilities, survival probabilities, and their first differences, by using parameter estimates of each model. Figure B.4 juxtaposes each model's estimates of the three events' conditional probabilities at y_T, provided that any event happens, $\Pr(y_E|D(y_E) = 1, y_T)$, as an example, in the case of $\theta_{TE} = -0.9$ and $x_1 = x_2 = 0$. In this and the following figures, the bold lines and the shaded areas continue to be the mean and 95% confidence interval of the COEHA estimates; the solid and dashed lines are the case of the LST-CR model; the points and the vertical lines are the case of the MNL-CR model; and the true values are not shown because, unsurprisingly, they are almost the same as the mean values of COEHA estimates (the true model).

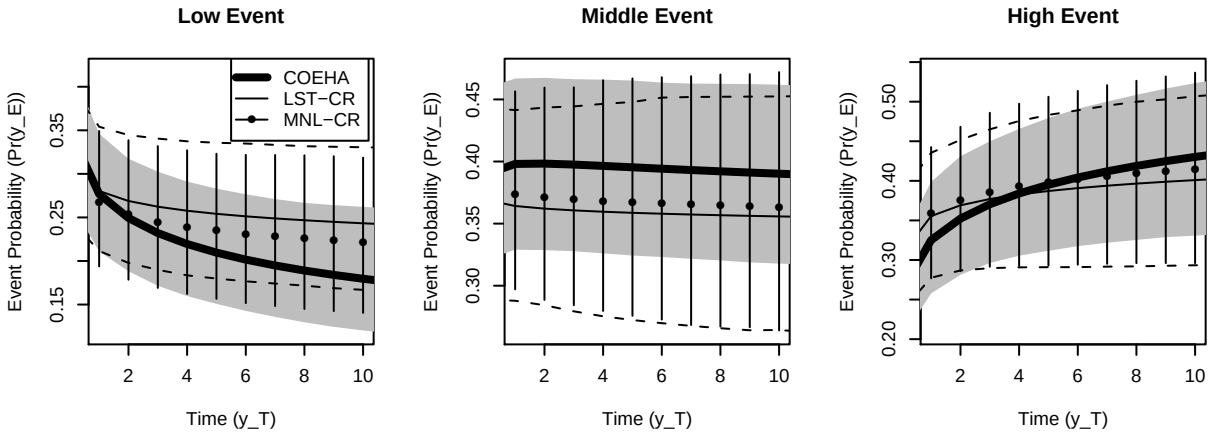

Figure B.4: Event probability against time y_T, conditional on one type of event happening, $\Pr(y_E|D(y_E) = 1, y_T, \theta_{TE} = -0.9, x_1 = x_2 = 0)$, in the case of regular censoring time $(z_T^0 = 50)$.

The two CR estimates are biased and less efficient. The LST-CR estimates of the Low (or Middle) event probability are upwardly (or downwardly) biased. In addition, the Low (or High) event probability estimated by the two CR models decreases (or increases) in time

more slowly than the true value. Moreover, the 95% confidence intervals of the two CR estimates are larger than those of the COEHA estimates.

First differences are also estimated with bias by the two CR models. In the right panel of Figure B.5, the vertical axis shows how much the High event probability at $y_T = 1$ changes if x_2 changes from 0 to 1, conditional on any event happening and $x_1 = 0$. The horizontal axis corresponds to the various values of θ_{TE}. The two CR first differences are downwardly biased due to *dependence among events*, which spoils an assumption of the two CR models, conditional independence among risks. (Note that the size of the bias does not depend on the value of θ_{TE}, the measure of dependence between time and events.) For instance, the unobserved (or unobservable) advantage of rebels (u_E) should make a rebel victory ($y_E = 2$) more likely and a government victory ($y_E = 0$) less likely. Thus, the larger $\Pr(y_E = 2)/\Pr(D(y_E) = 0)$, the smaller $\Pr(y_E = 0)/\Pr(D(y_E) = 0)$, which violates the independence of irrelevant alternatives on which the MNL-CR model is based. Or, the hazard of rebel victory is larger if the rebels win (larger u_E) than it would be if the government won (smaller u_E), which violates the non-informative censoring on which the LST-CR model is premised.

The left panel of Figure B.5 displays the first differences of x_1 on the High event at $y_T = 1$. The true first differences are positive (or negative) in the case of $\theta_{TE} < 0$ (or $\theta_{TE} > 0$), even though $\beta_{E1} = 0$. This is caused by dependence between time and events. If x_1 changes from 0 to 1, the hazard ($\exp(\boldsymbol{x}_T\boldsymbol{\beta}_T)^{\alpha_T}$) increases ($\because \beta_{T1} > 0$) and the distribution of Z_T shifts downward. Given that $y_T = z_T = 1$ is fixed, $u_T = F_T(z_T)$ becomes larger;

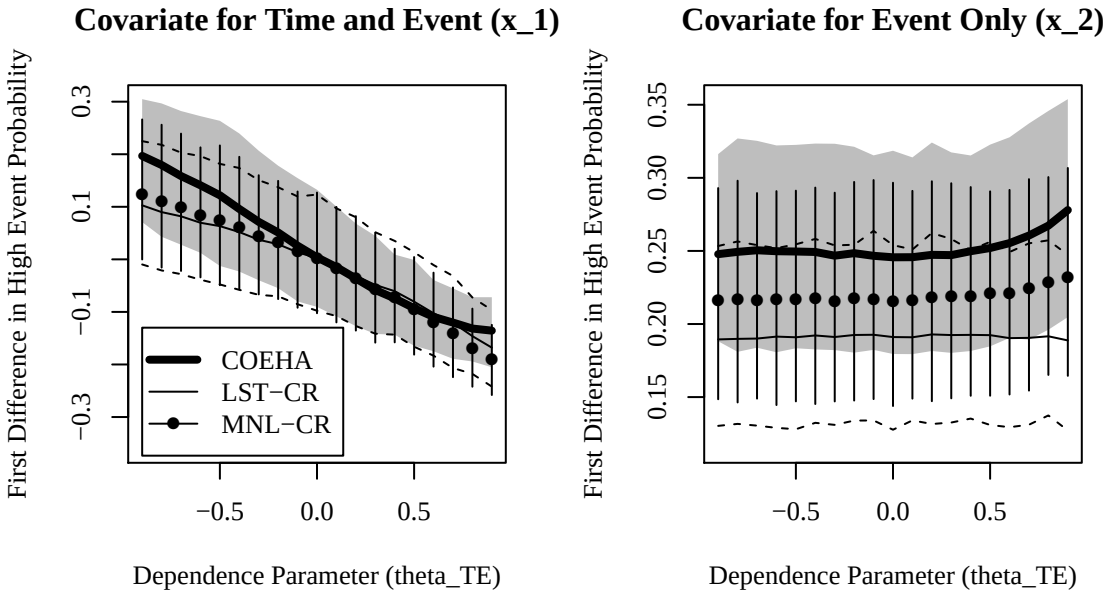

Figure B.5: First differences in High event probability at $y_T = 1$ against dependence parameter θ_{TE} in the case of regular censoring time ($z_T^0 = 50$). The vertical axis shows $\Pr(y_E = 2|D(y_E) = 1, y_T = 1, x_1 = 1, x_2 = 0) - \Pr(y_E = 2|D(y_E) = 1, y_T = 1, x_1 = 0, x_2 = 0)$ and $\Pr(y_E = 2|D(y_E) = 1, y_T = 1, x_1 = 0, x_2 = 1) - \Pr(y_E = 2|D(y_E) = 1, y_T = 1, x_1 = 0, x_2 = 0)$ in the left and right panels, respectively.

$F_T(z_T = 1|x_1 = 1) > F_T(z_T = 1|x_1 = 0)$. If there is positive dependence between U_T and U_E ($\theta_{TE} < 0$, confusingly, because of the *event-survival* AMH copula), the larger u_T, the larger u_E. Thus, the High event becomes more likely. For example, if we order the outcomes in a civil war as government victory ($y_E = 0$), settlement ($y_E = 1$), and rebel victory ($y_E = 2$), then when a civil war ends at $y_T = 1$, it is evaluated to be surprisingly late if rebels are strong (i.e. $x_1 = 1$, which should cause the conflict to end earlier), compared to a situation in which rebels are weak ($x_1 = 0$). Therefore, the rebel's victory ($y_E = 2$) is expected more if rebels are strong than if rebels are weak, even though the rebel's strength does not increase the probability of their victory for the whole period ($\beta_{E1} = 0$).

First differences of the two CR models are biased upward, especially if $\theta_{TE} < 0$, because the two CR models are not the true models and cannot capture *dependence between time and events* correctly. The size and direction of bias vary across the values of θ_{TE}, x_2, and y_T (whose results are not shown), while, in the case of $\theta_{TE} = 0$ (i.e. no dependence between time and events), estimates of all methods are unbiased.

Figure B.6 depicts the survival probability that any kind of event has not happened up to time y_T, $\overline{F}_T(y_T)$, in the case of $\theta_{TE} = -0.9, x_1 = x_2 = 0$. The two CR estimates are biased upward. The 95% confidence intervals of the MNL-CR estimates do not cover the true value, while those of the LST-CR estimates barely do so. Figure B.7 illustrates first differences in the survival probability at time $y_T = 1$ for various values of θ_{TE}. The left panel is the case where x_1 changes from 0 to 1 and $x_2 = 0$. The MNL-CR first differences are biased upward, while the LST-CR first differences are biased downward. The reason is that, as I elaborated with regard to why the CR coefficients of x_2 on events are biased, the hazards of the CR models are biased due to *dependence among events*. Note also that the value of θ_{TE} (the degree of dependence between time and events) does not matter for the size of the bias. In the right panel, x_2 moves from 0 to 1 ($x_1 = 0$). The COEHA first differences are exactly zero by design. The two CR first differences are a bit biased, though their 95% confidence intervals are mostly large enough to include the true value, zero.

APPENDIX C:

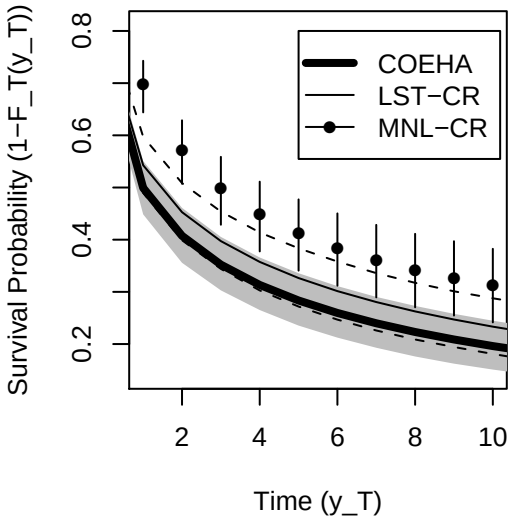

Figure B.6: Survival probability against time y_T, $\overline{F}_T(y_T|\theta_{TE} = -0.9, x_1 = x_2 = 0)$ in the case of regular censoring time ($z_T^0 = 50$).

ADDITIONAL ANALYSES OF APPLICATION

This appendix provides additional analyses of Cunningham et al. (2009)'s civil war data.

C.1 Combining Two Event Categories into One

If conditional independence among risks, an assumption of CR models, is correct, then the collapsing procedure (which combines formal agreement and low activity into a new category, "no victory") should not affect CR model estimates of the covariates' effects on the other events, government and rebel victories. In fact, when I apply the MNL-CR model to the three and four category event datasets, the coefficients on government and rebel victories do not change much (Tables C.1). The left two columns show the case of the original four

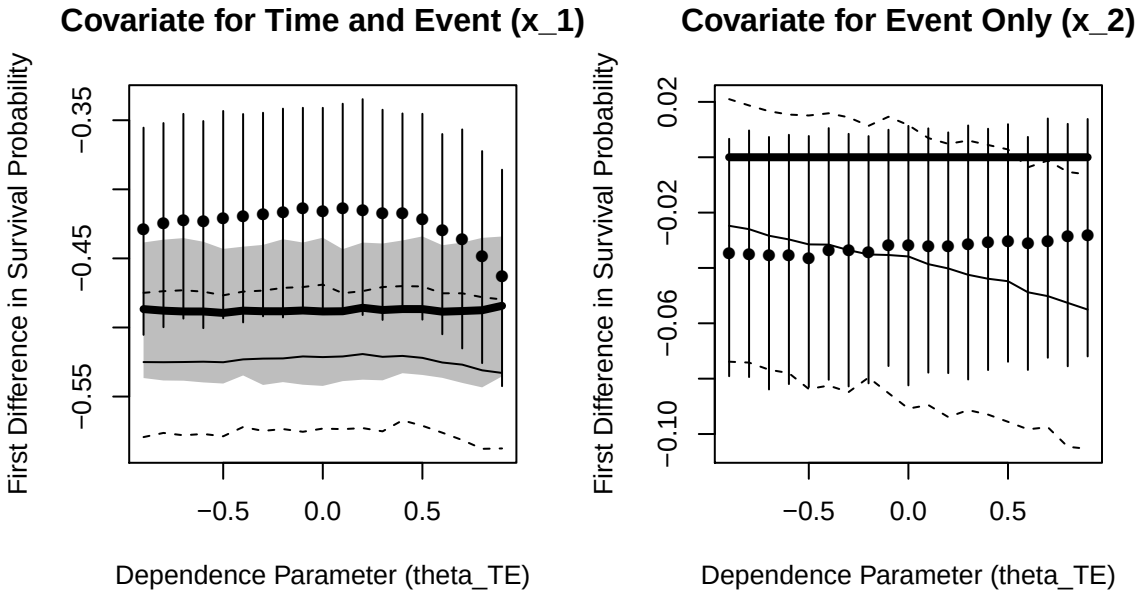

Figure B.7: First differences in survival probability at $y_T = 1$ against dependence parameter θ_{TE} in the case of regular censoring time ($z_T^0 = 50$). The vertical axis shows $\overline{F}_T(1|x_1 = 1, x_2 = 0) - \overline{F}_T(1|x_1 = 0, x_2 = 0)$ and $\overline{F}_T(1|x_1 = 0, x_2 = 1) - \overline{F}_T(1|x_1 = 0, x_2 = 0)$ in the left and right panel, respectively. The bold lines and the shaded areas are the mean and 95% confidence interval of COEHA estimates. The solid and dashed lines are the case of the LST-CR model. The points and vertical lines are the average and 95% confidence interval of the MNL-CR model.

category event dataset (the estimates of the effects on formal agreement and low activity outcomes are omitted), while the right two columns display the case of the three category event dataset (the estimates of the effects on no victory are hidden). Note that the first and third columns are similar, while the second and fourth columns are comparable. Therefore, it is safely said that this collapsing procedure should not be responsible for the results of this paper.

Moreover, in the literature on civil wars, three-category event variables are not uncommon (Mason et al. 1999; Balch-Lindsay et al. 2008; Thyne 2009). Brandt et al. (2008, fn. 2) also

Table C.1: Parameter estimates of the MNL-CR models of Cunningham et al. (2009)'s civil war data (1946-2003)

Dataset	Four Categories				Three Categories			
Victory Outcome	Government		Rebel		Government		Rebel	
Constant	−2.679		4.433		−2.621		4.417	
	(1.891)		(2.348)		(1.889)		(2.342)	
Time at State	−0.458	**	−0.313	**	−0.459	**	−0.315	**
	(0.064)		(0.088)		(0.064)		(0.088)	
Territorial Control	0.254		−0.300		0.251		−0.309	
	(0.313)		(0.392)		(0.314)		(0.392)	
Rebels Stronger	0.266		2.380	**	0.262		2.385	**
	(0.840)		(0.613)		(0.839)		(0.613)	
Rebels at Parity	0.018		1.279	**	0.059		1.310	**
	(0.530)		(0.453)		(0.528)		(0.453)	
Legal Political Wing	0.899	*	0.777		0.898	*	0.791	
	(0.374)		(0.493)		(0.375)		(0.492)	
ELF Index	−0.881		−1.052		−0.870		−1.032	
	(0.612)		(0.739)		(0.612)		(0.738)	
Ethnic Conflict	−0.482		−1.763	**	−0.476		−1.765	**
	(0.342)		(0.639)		(0.342)		(0.638)	
Ln GDP per Capita	0.198		−0.700	**	0.196		−0.698	**
	(0.173)		(0.228)		(0.173)		(0.228)	
Democracy	−2.280	**	−1.401		−2.270	**	−1.406	
	(0.633)		(0.766)		(0.633)		(0.767)	
Two or More Dyads	−0.571		0.182		−0.573		0.178	
	(0.321)		(0.383)		(0.321)		(0.382)	
Ln Population	0.177		−0.030		0.172		−0.028	
	(0.116)		(0.154)		(0.115)		(0.153)	

NOTE: ** $p < 0.01$, * $p < 0.05$. N $= 2,201$. The left two columns show the case of the original four category event dataset, while the right two columns display the case of the three category event datase. The cluster robust standard errors are in parenthesis.

"combined negotiated settlements and truce."

C.2 Exclusion Restriction

Since the same covariates are used for time ($\boldsymbol{x}_T$) and event ($\boldsymbol{x}_E$) regressors in the specification used in the manuscript (Table 1), some might worry about the exclusion restriction. To address this concern, I use two other specifications in which I include the following four other indicators of rebel's military strength which CGS (pp. 580-581 and their Table 3) used as regressors for time (the median is in parenthesis).

- *Strong Central Command*: a dummy variable for whether a rebel group has a clear leadership structure and whether the leadership exercises a high degree of control over their organization. (0)

- *High Mobilization Capacity*: a dummy variable for whether mobilization capacity is high relative to the government. (0)

- *High Arms-procurement Capacity*: a dummy variable for whether a rebel group has high arms procurement capacity relative to the government. (0)

- *High Fighting Capacity*: a dummy variable for whether a rebel group is deemed to have high fighting capacity relative to the government. (0)

I include these alternative indicators both in exchange of (Table C.2), and in addition to (Table C.3), the original indicators of rebel's military strength (i.e. Rebels Stronger and Rebels at Parity) as regressors for time. The coefficient of Rebels Stronger on time is significant in Table 1 but is not significant in Table C.3 probably because of multicollinearlity. Otherwise, however, the results do not change substantially.

Table C.2: Parameter estimates of the COEHA model of Cunningham et al. (2009)'s civil war data (1946-2003)

	Time (F_T)				Event (F_E)			
	Coef.		95% CI		Coef.		95% CI	
Territorial Control	−0.28	*	(−0.62,	−0.00)	−0.27		(−0.99,	0.43)
Strong Central Command	0.25		(−0.09,	0.59)				
High Mobilization Capacity	0.32		(−0.10,	0.89)				
High Arms Procurement Capacity	1.17		(−0.03,	2.88)				
High Fighting Capacity	0.37		(−0.65,	1.44)				
Rebels Stronger					1.60		(−0.48,	5.23)
Rebels at Parity					1.03		(−0.30,	2.54)
Legal Political Wing	0.49	*	(0.06,	0.96)	−0.07		(−0.81,	0.74)
ELF Index	0.27		(−0.32,	1.03)	0.47		(−0.97,	1.94)
Ethnic Conflict	−0.00		(−0.33,	0.38)	−0.14		(−0.79,	0.48)
Ln GDP per Capita	0.04		(−0.13,	0.24)	−0.31		(−0.64,	0.02)
Democracy	−0.93	**	(−1.40,	−0.39)	0.73	*	(0.09,	1.37)
Two or More Dyads	−0.41	**	(−0.68,	−0.14)	0.47		(−0.16,	1.25)
Ln Population	−0.05		(−0.19,	0.07)	−0.10		(−0.34,	0.13)
$\kappa(0)$					−4.41	**	(−7.92,	−1.38)
$\log(\kappa(1) - \kappa(0))$					1.20	**	(1.07,	1.43)
	Time and Event (F_{TE})							
$\log((1+\theta_{TE})/(1-\theta_{TE}))$	−13.04	**	(−13.40,	−6.96)				

NOTE: ** $p < 0.01$, * $p < 0.05$. N = 2,201. The Cox model 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.

In addition, the above Monte Carlo simulation demonstrates that, even though the only variable included as a regressor for time, x_1, is also included as a regressor for event, the parameter estimates are unbiased. All told, it is concluded that we do not have to care about the exclusion restriction.

Table C.3: Parameter estimates of the COEHA model of Cunningham et al. (2009)'s civil war data (1946-2003)

	Time (F_T)				Event (F_E)			
	Coef.		95% CI		Coef.		95% CI	
Territorial Control	−0.37	*	(−0.76,	−0.04)	−0.26		(−0.97,	0.45)
Strong Central Command	0.26		(−0.08,	0.62)				
High Mobilization Capacity	0.33		(−0.09,	0.89)				
High Arms Procurement Capacity	0.99		(−0.62,	2.55)				
High Fighting Capacity	0.09		(−1.33,	1.25)				
Rebels Stronger	0.48		(−0.18,	2.43)	1.46		(−0.63,	4.93)
Rebels at Parity	0.41		(−0.10,	0.93)	0.90		(−0.43,	2.44)
Legal Political Wing	0.51	*	(0.07,	1.01)	−0.06		(−0.80,	0.75)
ELF Index	0.23		(−0.37,	1.01)	0.47		(−0.97,	1.95)
Ethnic Conflict	0.02		(−0.30,	0.41)	−0.15		(−0.80,	0.46)
Ln GDP per Capita	0.04		(−0.14,	0.25)	−0.32		(−0.64,	0.02)
Democracy	−0.93	**	(−1.41,	−0.40)	0.72	*	(0.08,	1.37)
Two or More Dyads	−0.39	**	(−0.66,	−0.11)	0.46		(−0.16,	1.25)
Ln Population	−0.03		(−0.17,	0.08)	−0.11		(−0.35,	0.13)
$\kappa(0)$					−4.52	**	(−7.96,	−1.44)
$\log(\kappa(1)-\kappa(0))$					1.19	**	(1.07,	1.43)
$\log((1+\theta_{TE})/(1-\theta_{TE}))$	−12.29	**	(−13.36,	−1.48)				

NOTE: ** $p < 0.01$, * $p < 0.05$. N = 2,201. The Cox model 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.

C.3 Model Specification

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. Note that the Gaussian, FGM and Frank copulas satisfy

$$\overline{C}^{(T)}(u_T, u_E|\theta_{TE}) = \overline{C}^{(E)}(u_T, u_E|\theta_{TE}) = C(u_T, u_E| - \theta_{TE})$$

and, therefore, are radially symmetric

$$\overline{C}^{(TE)}(u_T, u_E|\theta_{TE}) = C(u_T, u_E|\theta_{TE}).$$

That is, their associated copulas can be expressed using just the Gaussian, FGM or Frank copulas, respectively. Thus, I do not estimate these associated copulas. Table C.4 compares the AIC values of the 30 models. It turns out that a combination of the ordered logit model and the event-survival AMH copula has the lowest AIC value: 3065.7. Therefore, I use this model specification in subsection 3.4 of the main text.

Table C.4: AIC values of the COEHA models of Cunningham et al. (2009)'s civil war data (1946-2003)

Event Model F_E	Copula C	Original C	Time Survival $\overline{C}^{(T)}$	Event Survival $\overline{C}^{(E)}$	Survival $\overline{C}^{(TE)}$
Ordered Logit	Gaussian	3068.1			
	FGM	3067.4			
	Frank	3067.4			
	AMH	3067.5	3068.4	3065.7	3067.8
	Clayton	3073.7	3073.7	3073.7	3073.7
	Gumbel	3073.7	3073.7	3073.7	3073.7
Ordered Probit	Gaussian	3068.2			
	FGM	3067.5			
	Frank	3067.4			
	AMH	3067.4	3068.4	3065.8	3067.9
	Clayton	3073.3	3073.3	3073.3	3073.3
	Gumbel	3073.3	3073.3	3073.3	3073.3

NOTE: N = 2,201. The Cox model for time (F_T), the ordered logit model for event (F_E), and various copulas for dependence between time and events (C) are employed.

References

Ali, M. M., Mikhail, N. N., and Haq, M. S. (1978), "A Class of Bivariate Distributions Including the Bivariate Logistic," *Journal of Multivariate Analysis*, 8, 405–412.

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.

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

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.

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.

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.

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.