

# A Bayesian Split Population Survival Model for Duration Data With Misclassified Failure Events

**Benjamin E. Bagozzi<sup>1</sup>, Minnie M. Joo<sup>2</sup>, Bomin Kim<sup>3</sup> and Bumba Mukherjee<sup>4</sup>**

<sup>1</sup> Assistant Professor, Dept. of Political Science & International Relations, University of Delaware, USA.

Email: [bagozzib@udel.edu](mailto:bagozzib@udel.edu)

<sup>2</sup> Assistant Professor, Dept. of Political Science, University of Massachusetts Lowell, USA. Email: [minhyung.joo@gmail.com](mailto:minhyung.joo@gmail.com)

<sup>3</sup> Ph.D. Candidate, Dept. of Statistics, Penn State, USA. Email: [bomin8319@gmail.com](mailto:bomin8319@gmail.com)

<sup>4</sup> Professor, Dept. of Political Science, Penn State, USA. Email: [sxm73@psu.edu](mailto:sxm73@psu.edu)

## Abstract

We develop a new Bayesian split population survival model for the analysis of survival data with misclassified event failures. Within political science survival data, right-censored survival cases are often erroneously misclassified as failure cases due to measurement error. Treating these cases as failure events within survival analyses will underestimate the duration of some events. This will bias coefficient estimates, especially in situations where such misclassification is associated with covariates of interest. Our split population survival estimator addresses this challenge by using a system of two equations to explicitly model the misclassification of failure events alongside a parametric survival process of interest. After deriving this model, we use Bayesian estimation via slice sampling to evaluate its performance with simulated data, and in several political science applications. We find that our proposed “misclassified failure” survival model allows researchers to accurately account for misclassified failure events within the contexts of civil war duration and democratic survival.

**Keywords:** Bayesian methods, duration models, mixture distributions

## 1 Introduction

Researchers often analyze data on the time until an event occurred (or “failed”), otherwise known as “survival data.” For Political Science survival data, one’s ability to record an event as having failed at a given point in time is frequently prone to measurement error. More precisely, Political Science survival data can often *over-report* events as having failed, such that some observations’ true *censored* values are misclassified as *failed*. This produces an unobservable mixture of failure events within one’s data, with some failure events corresponding to true failure events, and other failure events representing cases that have been incorrectly misclassified as failures (i.e., “misclassified failures”). Misclassified failures (MFs) are in actuality right-censored events: the researcher should only conclude that the observation lasted up until the recorded failure time. Concluding instead that the observation failed at that point in time is problematic as there is a non-zero probability that the observation persisted past that point.

There are several scenarios where (a subset of) recorded failure events may actually persist beyond their recorded failure time in this manner, leading to misclassification in event failures.

*Political Analysis* (2019)

DOI: 10.1017/pan.2019.6

**Corresponding author**

Benjamin E. Bagozzi

**Edited by**

Jeff Gill

© The Author(s) 2019. Published by Cambridge University Press on behalf of the Society for Political Methodology.

*Authors’ note:* Earlier versions of this paper were presented as a poster at the 2014 Society For Political Methodology Annual Meeting in Athens, GA, USA and the 2018 Asian Political Methodology Conference in Seoul, South Korea, and as a paper at the 2018 Penn State Department of Political Science Brown Bag seminar series in State College, PA, USA, and the 2018 Visions in Methodology Conference at Ohio State University, in Columbus, OH, USA. Bagozzi’s contribution is partly based upon the work supported by the National Science Foundation under Grant Nos. SBE-SMA-1539302 and DMS-1737865. The authors wish to thank Jeff Gill, two anonymous reviewers, Lee Ann Banaszak, Liz Carlson, Bruce Desmarais, Kentaro Fukumoto, Simon Heuberger, Nahomi Ichino, Kosuke Imai, Azusa Katagiri, Fridolin Linder, Elizabeth Menninga, Shawna K. Metzger, Sara Mitchell, Inken von Borzyskowski, Michael Ward, Joe Wright, and Teppei Yamamoto for their helpful comments and suggestions. See Bagozzi *et al.* (2019) for replication materials.

For example, in many Political Science applications one's events of interest often do not have clearly observable end points (i.e., "failures"). When this is the case, the researcher must establish a threshold criterion to determine whether (and when) a duration observation (or some subset of observations) failed. Often the strategy is to choose a failure threshold that, if anything, *underestimates* the length of one's actual event. The implicit reasoning for this is that it is better to be conservative and ensure that coded events end before they truly do than it is to code events as incorrectly persisting beyond their true failures. As an example, consider research on civil war duration. Here, researchers typically analyze the durations of rebel-government conflicts, but record civil war end dates ("failures") for specific conflicts based upon 24-month spells with fewer than 25 battle deaths per year (e.g., Balch-Lindsay and Enterline 2000, Buhaug, Gates, and Lujala 2009, Thyne 2012). This threshold is overly conservative, especially for lower-intensity civil wars in remote or poor information environments that persist indefinitely with little actual fighting.<sup>1</sup> Treating these cases as failures in survival analyses can lead to bias, especially if covariates of interest happen to be correlated with an observation's likelihood of misclassification of failure—as demonstrated in the sections further below.

Misclassified failure events can also arise in survival data due to a variety of other coding or reporting processes. For example, within long-range historical analyses, studies of the durations of ancient civilizations or political processes therein (e.g., Cioffi-Revilla and Lai 1995; Cioffi-Revilla and Landman 1999) typically do not have data on the precise time point of a given *failure* event due to the sands of time. Instead, researchers must make do with the best available proxy for such a failure event, often using the *last known* historical record (e.g., artifact or carbon dating) of an ancient civilization or social activity. Here, each observation's recorded failure time is an *underestimate* of that observation's true life span, in that a researcher knows with certainty that the observation lasted at least up until that point, but there is a strong likelihood that it persisted for some amount of time past that recorded failure. To the extent that these underestimates of duration are non-random, and are correlated with commonly studied covariates (e.g., environmental or geographic conditions), bias will again arise in survival estimates of these phenomena. Finally, political actors often self-report their duration of (non)engagement in a given political activity, and these reports are often leveraged within survival analyses (e.g., Cress, McPherson, and Rotolo 1997; Box-Steffensmeier, Radcliffe, and Bartels 2005). In some cases, these actors may strategically under-report their duration of (non)engagement, ensuring that the recorded failures in one's survival data to exhibit misclassification.

To address the methodological challenges associated with MFs, we develop a parametric MF survival model that explicitly accounts for the potential that an unknown subset of failure events actually "lived on" beyond a researcher's recorded failure times for those observations. Our proposed model does so by estimating a system of two equations. The first is a "splitting" equation that estimates the probability of a case being a misclassified failure, with or without covariates. The second equation then represents that of a standard parametric survival model, whose relevant failure and survival probabilities are now estimated conditional on a case *not* being a misclassified failure.

As such, our model shares similarities with the cure survival model, which has been previously used in Political Science to model competing processes of democratic survival (Svolik 2008), or to accommodate heterogeneous mixtures of "at risk" and "not at risk" countries in global analyses of irregular leadership changes (Beger, Dorff, and Ward 2014, 2015; Beger *et al.* 2017). However, in contrast to the cure model—which only allows one to model heterogeneous mixtures among observations that have not failed—our proposed model allows one to specifically account for

1 A similar case is the survival of terrorist groups, where scholars code and analyze survival as the time between a terrorist group's first and last known attacks (Young and Dugan 2014).

heterogeneous mixtures of failure cases. This in turn suggests that both models can alternatively be seen as “inflated” models, wherein our proposed model accounts for inflation in failure cases and the cure model accounts for inflation in non-failure (i.e., right-censored) cases. In this light, our paper also contributes to researchers’ broader efforts to disentangle mixtures of multiple survival data processes within the contexts of repeated events (Box-Steffensmeier, De Boef, and Joyce 2007), survival phases (Metzger and Jones 2016), and non-proportional hazards (Keele 2010; Licht 2011; Jin and Boehmke 2017; Ruhe 2018).

The remainder of this paper proceeds as follows. After deriving our model in non-time-varying and time-varying covariate contexts, we develop an R package to facilitate its estimation via Bayesian inference with a slice-sampling algorithm (i.e., a Markov Chain Monte Carlo method). We then illustrate the advantages of our Bayesian model within a series of Monte Carlo (MC) simulations and two separate political science applications. Notably, these illustrations reveal that our proposed model not only is capable of providing improved survival estimates—and theoretical insights—concerning the determinants of survival processes when MFs are present, but also offers researchers a means of theoretically identifying (and testing for) the factors that govern a particular MF process.

## 2 Survival Model with Misclassified Failure

### 2.1 Parametric Misclassified Failure Model

We formally describe below our new split population survival model—labeled as the “misclassified failure” (MF) model—that explicitly models the misclassification probability of failure versus right-censored events. We first define our MF model’s general parametric log-likelihood function, which can be used in conjunction with commonly used parametric survival models (e.g., exponential, Weibull, or log-normal). We then use this general MF framework to develop our main model of interest—the Bayesian MF *Weibull* model with time-varying covariates—that is estimated by Markov Chain Monte Carlo (MCMC) methods.

We start by defining a general parametric survival model for continuous time duration data, where subjects  $i = \{1, 2, \dots, N\}$  each eventually experience an event of interest. However, not all subjects need experience the event during a particular sample period, as some may survive until the end of the sampling window, in which case they are “censored” in their final period of observation ( $\tilde{C}_i = 0$  if censored, and 1 otherwise). The duration of interest  $t$  is thus assumed to have a probability density function (PDF) of  $f(t) = \Pr(T_i = t)$ , where  $T$  is an observation’s duration of time until experiencing the event or censoring. The cumulative distribution function (CDF) for the probability of the event on or before  $t$  is accordingly  $\Pr(T_i \leq t) \equiv F(t) = \int_0^t f(t) dt$ , where the probability of survival is  $\Pr(T_i \geq t) \equiv S(t) = 1 - F(t)$ . With this PDF and CDF, the hazard of an event at  $t$  given that the event has not occurred prior to that point is  $h(t) = \frac{f(t)}{S(t)}$ . We next use these probability statements to define the (log) likelihood for a general parametric survival model.

To this end, note that uncensored observations ( $\tilde{C}_i = 1$ ) provide information on both the hazard of an event and the survival of individuals prior to that event, whereas censored observations ( $\tilde{C}_i = 0$ ) only provide information on an observation having survived at least until time  $T_i$ . Combining each set of observation’s respective contributions to the density and survival functions, the likelihood and the log-likelihood function(s) of the standard parametric survival model are respectively,

$$L = \prod_{i=1}^N [f(t_i)]^{\tilde{C}_i} [S(t_i)]^{1-\tilde{C}_i} \quad \text{and} \quad (1)$$

$$\ln L = \sum_{i=1}^N \{\tilde{C}_i \ln[f(t_i|\mathbf{X}_i, \boldsymbol{\beta})] + (1 - \tilde{C}_i) \ln[S(t_i|\mathbf{X}_i, \boldsymbol{\beta})]\}, \quad (2)$$

where  $\mathbf{X}_i$  are  $p_1$ -dimensional covariates and  $\boldsymbol{\beta}$  is the corresponding parameter vector in  $\mathbb{R}^{p_1}$ . We build on this standard survival model to account for asymmetric misclassification arising within one's censored and failure observations to develop our MF model. To do so, we focus on situations where censored cases are misclassified as *failed* observations, in which case one's observed censoring indicator  $\tilde{C}_i$  accurately records all censored cases ( $\forall (\tilde{C}_i = 0) : (C_i = 0)$ ) but mis-records some subset of non-censored failure outcomes as censored ( $\exists (\tilde{C}_i = 1) : (C_i = 0)$ ). Drawing on Box-Steffensmeier and Zorn's (1999) notation in their review of the cure survival model, we define a corresponding probability of misclassification as  $\alpha_i = \Pr(\tilde{C}_i = 1 | C_i = 0)$ . This implies that the unconditional density is defined by the combination of an observation's misclassification probability and its probability of experiencing an actual failure conditional on not being misclassified:

$$\Pr(\tilde{C}_i = 1 | C_i = 0) + \Pr(\tilde{C}_i = 0 | C_i = 0) \Pr(t_i \leq T_i) = \alpha_i + (1 - \alpha_i)f(t_i), \quad (3)$$

with the corresponding unconditional survival function of

$$\Pr(\tilde{C}_i = 0 | C_i = 0) \Pr(t_i > T_i) = (1 - \alpha_i)S(t_i), \quad (4)$$

where  $\alpha_i$  can be estimated via a binary response function such as probit, complementary log–log, or logit and is thus defined for the logit case as:

$$\alpha_i = \frac{\exp(\mathbf{Z}_i \boldsymbol{\gamma})}{1 + \exp(\mathbf{Z}_i \boldsymbol{\gamma})}, \quad (5)$$

where  $\mathbf{Z}_i$  are  $p_2$ -dimensional covariates.  $\boldsymbol{\gamma}$  is the corresponding parameter vector in  $\mathbb{R}^{p_2}$ . Combining each set of observation's respective contributions to the density and survival functions, and given the expression for  $\alpha_i$  in (5), the log-likelihood function of the general parametric split population model with MF cases (*without* time-varying covariates) is

$$\ln L = \sum_{i=1}^N \{\tilde{C}_i \ln[\alpha_i + (1 - \alpha_i)f(t_i | \mathbf{X}_i, \boldsymbol{\beta})] + (1 - \tilde{C}_i) \ln[(1 - \alpha_i)S(t_i | \mathbf{X}_i, \boldsymbol{\beta})]\}. \quad (6)$$

We next extend our MF model developed above and the model's log-likelihood in (6) to account for time-varying covariates. To do so, we re-define our survival data with unique "entry time" duration  $t_0$  and "exit time" duration  $t$  for each period at which an observation is observed. As such,  $t_{0ij}$  denotes observation  $i$ 's elapsed time since inception until the beginning of time period  $j$  and  $t_{ij}$  denotes the elapsed time since that observation's inception until the end of period  $j$ . An observation's status at time  $t_{ij}$  is then coded as censored ( $\tilde{C}_{ij} = 0$ ) or as having failed or "ended" ( $\tilde{C}_{ij} = 1$ ) at time  $t_{ij}$ . For  $t$ , the PDF ( $f(t)$ ), CDF ( $F(t)$ ), probability of survival ( $S(t)$ ), and hazard of an event ( $h(t)$ ) remain as defined above. However, we must now also define the probability of survival up until period  $j$ , as

$$S(t_0) = 1 - F(t_0), \quad (7)$$

where  $F(t_0) = \int_0^{t_0} f(t) dt$ . With  $S(t_0)$  defined, we extend the general parametric survival model's log-likelihood defined in equation (2) to accommodate time-varying covariates  $\mathbf{X}_{ij}$  and associated parameter vectors of  $\boldsymbol{\beta}$  by conditioning an observation's hazard and survival probability for time

*t* upon its probability of survival until *t*0:

$$\ln L = \sum_{i=1}^N \left\{ \tilde{C}_{ij} \ln \left[ \frac{f(t_{ij} | \mathbf{X}_{ij}, \boldsymbol{\beta})}{S(t0_{ij} | \mathbf{X}_{ij}, \boldsymbol{\beta})} \right] + (1 - \tilde{C}_{ij}) \ln \left[ \frac{S(t_{ij} | \mathbf{X}_{ij}, \boldsymbol{\beta})}{S(t0_{ij} | \mathbf{X}_{ij}, \boldsymbol{\beta})} \right] \right\}. \quad (8)$$

As described in the Supplementary Appendix, we use the steps described in equations (3) to (6) and extend the log-likelihood function in (8) to define the log-likelihood function of the parametric MF model with time-varying covariates as:

$$\ln L = \sum_{i=1}^N \left\{ \tilde{C}_{ij} \ln \left[ \alpha_{ij} + (1 - \alpha_{ij}) \frac{f(t_{ij} | \mathbf{X}_{ij}, \boldsymbol{\beta})}{S(t0_{ij} | \mathbf{X}_{ij}, \boldsymbol{\beta})} \right] + (1 - \tilde{C}_{ij}) \ln \left[ (1 - \alpha_{ij}) \frac{S(t_{ij} | \mathbf{X}_{ij}, \boldsymbol{\beta})}{S(t0_{ij} | \mathbf{X}_{ij}, \boldsymbol{\beta})} \right] \right\} \quad (9)$$

where  $\alpha_{ij} = \frac{\exp(\mathbf{Z}_{ij} \boldsymbol{\gamma})}{1 + \exp(\mathbf{Z}_{ij} \boldsymbol{\gamma})}$  can be accordingly estimated via a logit CDF, or alternatively via a probit or a complimentary log-log CDF. Thus, as shown in the log-likelihood in (9), the MF model with time-varying covariates accounts for the probability of misclassification via  $\alpha_{ij}$  since the observed event failures may include latent MF cases *and* the influence of covariates on the hazard of the event of interest. Note that the general properties of the standard cure model also—as presented in Box-Steffensmeier and Zorn (1999, 5) and as shown in the Supplementary Appendix—hold for the MF model, including (i) the reduction of the latter to a standard parametric model when  $\alpha_{ij} = 0$  and (ii) parameter identification even in the case where identical covariates are included in  $\mathbf{Z}$  and  $\mathbf{X}$ . But in contrast to the standard cure model (which accounts for an excess number of subjects who are immune to experiencing an event), the MF model with time-varying covariates is a model for instances where some subjects are observed as having failed or experienced the event, even though they in actuality “live on” past their observed-failure point. Hence, the MF model is useful in situations where observed event *failures* in the survival data are contaminated with latent MF cases.<sup>2</sup>

The log-likelihood statement of the time-varying MF model in (9) can be used in conjunction with commonly used parametric survival models such as the exponential, Weibull, log-logistic, or log-normal. Since Political Science survival model applications that use duration data which are prone to the contamination of latent MF cases (e.g., civil conflict duration data) use the standard Weibull model, we develop and define the log-likelihood function of the MF *Weibull* model with time-varying covariates in the next section. In the Supplementary Appendix, we also develop and assess the MF *exponential* model via MC simulations.

## 2.2 Misclassified Failure Weibull Model

Suppose that the survival time *t* has a Weibull distribution of  $W(t_{ij} | \rho, X_{ij}, \boldsymbol{\beta})$ . Then the density function and survival function in this case are as follows:

$$\begin{aligned} f(t_{ij} | \rho, \mathbf{X}_{ij}, \boldsymbol{\beta}) &= \exp(\mathbf{X}_{ij} \boldsymbol{\beta})^\rho \rho (\exp(\mathbf{X}_{ij} \boldsymbol{\beta}) t_{ij})^{\rho-1} \exp(-\exp(\mathbf{X}_{ij} \boldsymbol{\beta}) t_{ij})^\rho \\ S(t_{ij} | \rho, \mathbf{X}_{ij}, \boldsymbol{\beta}) &= \exp(-(\exp(\mathbf{X}_{ij} \boldsymbol{\beta}) t_{ij})^\rho). \end{aligned} \quad (10)$$

In the Supplementary Appendix, we follow the steps in equations (3) to (6) *and* use the parametric time-varying MF model’s log-likelihood function in (9) to develop the log-likelihood function of the

2 Although our parametric split population survival model focuses on “misclassified failure” cases, scholars may (owing to coding error) also overlook cases that have truly failed and instead treat these “misclassified survival” cases as having survived past their true failed date. Developing a split population duration model that addresses this issue of “misclassified survival” requires further research. However, the model developed here can potentially be adjusted to develop a split population model that accounts for “misclassified survival” cases in survival datasets.

MF Weibull model with time-varying covariates, which is given by:

$$\begin{aligned} \ln L(\rho, \beta, \gamma) = & \sum_{i=1}^N \left\{ \tilde{C}_{ij} \ln \left[ \alpha_{ij} + (1 - \alpha_{ij}) \frac{\exp(\mathbf{X}_{ij}\beta)^\rho \rho (\exp(\mathbf{X}_{ij}\beta)t_{ij})^{\rho-1} \exp(-\exp(\mathbf{X}_{ij}\beta)t_{ij})^\rho}{\exp(-(\exp(\mathbf{X}_{ij}\beta)t_{0ij})^\rho)} \right] \right. \\ & \left. + (1 - \tilde{C}_{ij}) \ln \left[ (1 - \alpha_{ij}) \frac{\exp(-(\exp(\mathbf{X}_{ij}\beta)t_{ij})^\rho)}{\exp(-(\exp(\mathbf{X}_{ij}\beta)t_{0ij})^\rho)} \right] \right\}. \end{aligned} \quad (11)$$

The model's log-likelihood in (11) thus accounts for the probability of misclassification and covariates that influence the survival of the event of interest given by a Weibull distribution.

While the MF Weibull model with time-varying covariates can be estimated by maximum likelihood using, for example, BFGS,<sup>3</sup> we estimate this model via a MCMC algorithm employed for Bayesian inference. We adopt the Bayesian estimation framework due to its flexibility and the fact that it makes use of all available information and produces clear and direct inferences. We thus label our model as the Bayesian MF Weibull model given the use of MCMC estimation. To conduct Bayesian inference, we need to assign a prior for each of the MF Weibull model's three parameters— $\rho$ ,  $\beta$ , and  $\gamma$ —and then define the conditional posterior distribution of these parameters. Following standard practice (e.g., Carlin and Louis 2000), we assign the multivariate normal prior to  $\beta = \{\beta_1, \dots, \beta_{p_1}\}$  and  $\gamma = \{\gamma_1, \dots, \gamma_{p_2}\}$ , and the Gamma prior for  $\rho$  with shape and scale parameters  $a_\rho$  and  $b_\rho$ :

$$\begin{aligned} \rho &\sim \text{Gamma}(a_\rho, b_\rho), \quad \beta \sim \text{MVN}_{p_1}(\mathbf{0}, \Sigma_\beta), \quad \gamma \sim \text{MVN}_{p_2}(\mathbf{0}, \Sigma_\gamma) \\ \Sigma_\beta &\sim \text{IW}(S_\beta, v_\beta) \quad \Sigma_\gamma \sim \text{IW}(S_\gamma, v_\gamma), \end{aligned} \quad (12)$$

where  $a_\rho, b_\rho, S_\beta, v_\beta, S_\gamma, v_\gamma$  are the hyperparameters. The multivariate normal prior that we employ for estimation is a weakly informative prior. While we primarily use the multivariate normal prior for our analysis, we also evaluate the robustness of the MC simulation and empirical application results from our Bayesian MF Weibull model by separately assigning another weakly informative prior for estimation of this model, namely the multivariate Cauchy prior to  $\beta = \{\beta_1, \dots, \beta_{p_1}\}$  and  $\gamma = \{\gamma_1, \dots, \gamma_{p_2}\}$ , and the Gamma prior for  $\rho$  with shape and scale parameters  $a_\rho$  and  $b_\rho$  as described in (12) (where, as before,  $a_\rho, b_\rho, S_\beta, v_\beta, S_\gamma, v_\gamma$  are the hyperparameters).

We use Bayesian hierarchical modeling to estimate  $\Sigma_\beta$  and  $\Sigma_\gamma$  in (12) using the Inverse-Wishart (IW) distribution when using the multivariate normal or Cauchy prior. Given these prior specifications and the hyperparameters, the conditional posterior distributions for  $\rho$ ,  $\beta$ , and  $\gamma$  parameters in the Bayesian MF Weibull model (with time-varying covariates) are

$$\begin{aligned} P(\rho|\mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \mathbf{t0}, \beta, \gamma) &\propto P(\mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \mathbf{t0}, \beta, \gamma, \rho) \times P(\rho|a_\rho, b_\rho) \\ P(\beta|\mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \mathbf{t0}, \gamma, \rho) &\propto P(\mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \mathbf{t0}, \beta, \gamma, \rho) \times P(\beta|\Sigma_\beta) \\ P(\gamma|\mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \mathbf{t0}, \beta, \rho) &\propto P(\mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \mathbf{t0}, \beta, \gamma, \rho) \times P(\gamma|\Sigma_\gamma), \end{aligned}$$

where  $P(\mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \mathbf{t0}, \beta, \gamma, \rho)$  is the likelihood that can be obtained using the log-likelihood in equation (11), and  $P(\rho|a_\rho, b_\rho)$ ,  $P(\beta|\Sigma_\beta)$ , and  $P(\gamma|\Sigma_\gamma)$  are the priors in equation (12).

We next describe the sampling scheme used for our Bayesian inference. Because closed forms for the posterior distributions of  $\rho$ ,  $\beta$ , and  $\gamma$  are not available, we use MCMC methods with the following slice-sampling (Neal 2003) update scheme,

- Step 0. Choose initial value of  $\beta$ ,  $\gamma$ , and  $\rho$  and set  $i = 0$ .

<sup>3</sup> The Broyden, Fletcher, Goldfarb, Shannon (BFGS) method. In our MC analysis, we briefly assess the properties of the MF Weibull model estimated by BFGS.

- Step 1. Update  $\Sigma_\beta \sim P(\Sigma_\beta | \beta)$  and  $\Sigma_\gamma \sim P(\Sigma_\gamma | \gamma)$  from conjugate posteriors.<sup>4</sup>
- Step 2. Update  $\beta \sim P(\beta | \mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \mathbf{t0}, \gamma, \rho, \Sigma_\beta)$ ,  $\gamma \sim P(\gamma | \mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \mathbf{t0}, \beta, \rho, \Sigma_\gamma)$  and  $\rho \sim P(\rho | \mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \mathbf{t0}, \beta, \gamma, a_\rho, b_\rho)$  using slice sampling. We use the univariate slice sampler with stepout and shrinkage (Neal, 2003). Detailed steps to perform slice sampling for  $\beta$ ,  $\gamma$ , and  $\rho$  are described in the Supplementary Appendix.
- Step 3. Repeat Step 1 and Step 2 until the chain converges.
- Step 4. After  $N$  iterations, summarize the parameter estimates using posterior samples.

### 3 Monte Carlo Simulations

We conduct 15 MC experiments to compare the performance of the survival models discussed above. Each experiment evaluates samples of  $N = 1,000$ ,  $N = 1,500$ , and  $N = 2,000$ . Our primary MC experiments simulate either a non-MF Weibull-distributed outcome variable (Experiment 1) or a MF Weibull-distributed outcome variable (Experiment 2) and in each case compare the performance of a Bayesian Weibull model to that of a Bayesian MF Weibull model. Experiments 3–4 assess maximum likelihood estimated (via BFGS) Weibull and MF Weibull models for the same non-MF Weibull (Experiment 3) and MF Weibull (Experiment 4) simulated outcome variables. Experiments 5–8 simulate an exponentially distributed<sup>5</sup> outcome variable (Experiments 5 and 7), or a MF exponential outcome variable (Experiments 6 and 8), and compare the performance of (i) Bayesian Weibull, MF exponential, and MF Weibull models (Experiments 5–6) or (ii) BFGS exponential, Weibull, MF exponential and MF Weibull models (Experiments 7–8). Experiments 9–11 return to our Bayesian Weibull and Bayesian MF Weibull models and compare these two estimators under circumstances of increasingly larger MF rates. Finally, Experiments 12–15 reevaluate our primary MC results when using (i) an alternate prior specification or (ii) a (MF) log-logistic data generating process (d.g.p.).

For all experiments, we set  $sims = 500$  and assign our survival stage covariates ( $\mathbf{x}$ ) as  $\mathbf{x} = (\mathbf{1}, \mathbf{x}_1)'$  where  $\mathbf{x}_1$  is drawn from  $Uniform[-2.5, 12]$ . The primary MF experiments (Experiments 2, 4, 6, 8, 13, and 15) then add a moderate level of MF cases ( $\alpha = 5\%$ ) within the resultant survival outcome variable (Experiments 2, 4, 6, and 8), or add MF rates of 8%, 12%, and 15% (Experiments 9, 10, and 11). These modest levels of misclassification are anticipated to be comparable to MF rates within applications considered below. To generate MF rates, we define a set of misclassification stage covariates  $\mathbf{z} = (\mathbf{1}, \mathbf{z}_1, \mathbf{z}_2)'$ , where  $\mathbf{z}_1 = InUniform[0, 100]$  and  $\mathbf{z}_2 \equiv \mathbf{x}_1$ . Parameter values are assigned as  $(\beta_1, \beta_2)' = (1, 3.5)'$  for our survival stage predictors. Our misclassification stage parameters are defined as  $(\gamma_1, \gamma_2, \gamma_3)' = (-2, 3, 3)'$  (Experiments 2, 4, 6, 8, 13, and 15), or as  $(\gamma_1, \gamma_2, \gamma_3)' = (2, 1, 4)'$  (Experiment 9),  $(\gamma_1, \gamma_2, \gamma_3)' = (-3, 2, 5)'$  (Experiment 10), or  $(\gamma_1, \gamma_2, \gamma_3)' = (4.5, -1, 5)'$  (Experiment 11). The (MF) Weibull-distributed outcome variables (Experiments 1–4; 9–11, 13) use  $\rho = 2$ . The models used in Experiments 1–2, 4–6, 9–11, and 14–15 use a multivariate normal (weakly informative) prior; whereas the models employed in Experiments 12–13 use a multivariate Cauchy (very weakly informative) prior. For each parameter estimate, we retain and evaluate the (MCMC-simulated) mean estimate and standard error (MCSE), as well as the root mean square error (RMSE) and 95% (credible/confidence) coverage probabilities (CPs).

Experiment 1 evaluates the performance of (i) a Bayesian Weibull model and (ii) a Bayesian MF Weibull model when the true d.g.p. is Weibull with no MFs. We report these results in the top portion of Table 1, and also plot the full distributions of each model-specific parameter estimate within Figure A.1 of the Supplementary Appendix. In cases where a researcher encounters a non-MF Weibull-distributed outcome variable, we find in Table 1 and Figure A.1 that the Bayesian MF Weibull estimator exhibits comparable performance to a standard Bayesian Weibull model. For example, across all  $\beta$  parameters of interest, the Bayesian Weibull and Bayesian MF Weibull

4 The closed form of the full conditional distributions for  $\Sigma_\beta$  and  $\Sigma_\gamma$  are derived in the Supplementary Appendix.

5 A Weibull-distributed outcome variable with  $\rho = 1$ .

**Table 1.** Markov Chain Monte Carlo (MCMC)  $\beta$ —Estimates for Experiments 1 and 2.

Experiment 1: Non-MF Weibull D.G.P.									
#Obs.	Model	$\hat{\beta}_0$	MCSE( $\hat{\beta}_0$ )	RMSE( $\hat{\beta}_0$ )	CP( $\hat{\beta}_0$ )	$\hat{\beta}_1$	MCSE( $\hat{\beta}_1$ )	RMSE( $\hat{\beta}_1$ )	CP( $\hat{\beta}_1$ )
1000	Bayes Weibull	1.000	0.004	0.026	0.954	3.500	0.001	0.004	0.950
	Bayes MF Weibull	1.000	0.033	0.028	0.916	3.500	0.005	0.004	0.938
1500	Bayes Weibull	1.003	0.004	0.022	0.936	3.500	0.001	0.003	0.934
	Bayes MF Weibull	1.000	0.027	0.022	0.950	3.500	0.004	0.003	0.956
2000	Bayes Weibull	1.001	0.003	0.018	0.950	3.500	0.000	0.003	0.940
	Bayes MF Weibull	0.998	0.023	0.020	0.932	3.500	0.003	0.003	0.942
Experiment 2: MF Weibull D.G.P.									
#Obs.	Model	$\hat{\beta}_0$	MCSE( $\hat{\beta}_0$ )	RMSE( $\hat{\beta}_0$ )	CP( $\hat{\beta}_0$ )	$\hat{\beta}_1$	MCSE( $\hat{\beta}_1$ )	RMSE( $\hat{\beta}_1$ )	CP( $\hat{\beta}_1$ )
1000	Bayes Weibull	1.209	0.005	0.209	0.000	3.480	0.001	0.020	0.062
	Bayes MF Weibull	1.002	0.003	0.027	0.940	3.500	0.000	0.004	0.958
1500	Bayes Weibull	1.215	0.004	0.215	0.000	3.479	0.001	0.021	0.008
	Bayes MF Weibull	1.004	0.003	0.023	0.934	3.500	0.000	0.003	0.940
2000	Bayes Weibull	1.194	0.004	0.194	0.000	3.482	0.001	0.018	0.006
	Bayes MF Weibull	1.005	0.002	0.019	0.954	3.499	0.000	0.003	0.940

*Note:* True parameter values are  $\beta_0 = 1$  and  $\beta_1 = 3.5$ .

models recover averaged parameter estimates that are virtually identical. This is corroborated by the RMSEs and CPs reported in Table 1, which indicate that our Bayesian Weibull and Bayesian MF Weibull models recover  $\hat{\beta}$ 's with comparably low levels of bias, and comparably high levels of coverage, respectively. Indeed, there are several instances where the Bayesian MF Weibull model exhibits slightly less bias than the Bayesian Weibull. However, although both models consistently exhibit low MCSEs, the Bayesian Weibull model's MCSEs are consistently smaller than those of the Bayesian MF Weibull.

In sum, while the Bayesian Weibull model outperforms the Bayesian MF Weibull model in terms of efficiency, Experiment 1 suggests that (mis)applying the Bayesian MF Weibull to non-MF Weibull-distributed survival data does not lead to substantial biases in one's resulting parameter estimates. These conclusions are reinforced by Figure A.1, which demonstrates that the MF Bayesian Weibull model exhibits comparable parameter-estimate distributions (across 500 *sims*) to those of the standard Bayesian Weibull model, for all  $N$ 's considered.

Experiment 2 (re)evaluates the performance of the Bayesian Weibull and Bayesian MF Weibull models when the true d.g.p. is MF Weibull. We report these MC results in the lower half of Table 1 ( $\beta$  parameters) and in Table 2 ( $\gamma$  parameters). We also plot the full distributions of each  $\beta$  parameter in the Supplementary Appendix. These tables and figures reveal very favorable results for the Bayesian MF Weibull model, and less than favorable results for the Bayesian Weibull model. Looking first at the  $\beta$  estimates reported in Table 1, the Bayesian MF Weibull  $\hat{\beta}$ 's are highly comparable to our true parameter values, and improve in this respect as the number of observations is increased from 1,000 to 2,000. In fact, as the sample size increases, the obtained values of the beta covariates not only converge to their true theoretical value but also the RMSEs of these values shrink further to negligible levels and the 95% Empirical CPs remain above 90% for these obtained values. This broadly suggests that as the size of the dataset increases to even modest levels, our (Bayesian) MF models provide reliable and accurate estimates. By contrast, the standard Bayesian Weibull model's mean  $\hat{\beta}$ 's substantially overestimate  $\beta_0$  and typically underestimate  $\beta_1$  no matter the  $N$  considered. These conclusions are reinforced by the RMSE and CP values reported in Table 1. Herein, the Bayesian MF Weibull exhibits (i) RMSEs that

**Table 2.** Markov Chain Monte Carlo (MCMC)  $\gamma$ —Estimates for Experiment 2.

Experiment 2: MF Weibull D.G.P.													
#Obs.	Model	$\hat{\gamma}_0$	MCSE( $\hat{\gamma}_0$ )	RMSE( $\hat{\gamma}_0$ )	CP( $\hat{\gamma}_0$ )	$\hat{\gamma}_1$	MCSE( $\hat{\gamma}_1$ )	RMSE( $\hat{\gamma}_1$ )	CP( $\hat{\gamma}_1$ )	$\hat{\gamma}_2$	MCSE( $\hat{\gamma}_2$ )	RMSE( $\hat{\gamma}_2$ )	CP( $\hat{\gamma}_2$ )
1000	Bayes MF Weibull	-1.314	0.226	0.975	0.880	2.020	0.127	0.347	0.908	3.395	0.141	0.545	0.922
1500	Bayes MF Weibull	-1.150	0.185	0.992	0.802	1.962	0.096	0.294	0.882	3.374	0.108	0.485	0.880
2000	Bayes MF Weibull	-1.362	0.228	0.879	0.828	2.041	0.133	0.313	0.904	3.417	0.141	0.505	0.870

Note: True parameter values are  $\gamma_0 = -2$ ,  $\gamma_1 = 2$ , and  $\gamma_2 = 3$ .

are 5–10 times smaller than the Bayesian Weibull models’ RMSEs and (ii) CPs of 93%–95%—far higher than the Bayesian Weibull’s CPs (of 0%–8%).

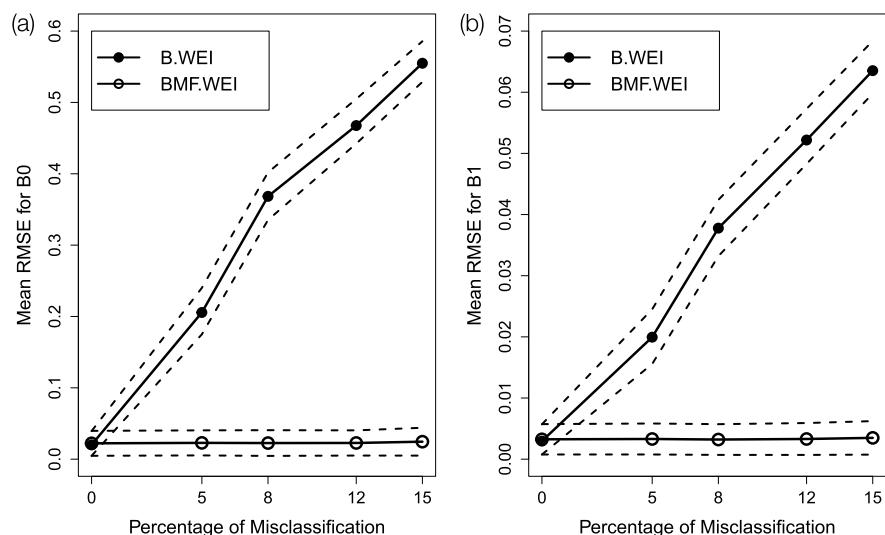
The full parameter distributions presented in Figure A.2 reinforce the above observations in demonstrating that—relative to the Bayesian MF Weibull model—the Bayesian Weibull’s  $\hat{\beta}$ ’s do a substantially worse job in capturing the true parameter values, across all sets of 500 simulations examined in Experiment 2. Turning to Experiment 2’s MF Weibull  $\gamma$  estimates (Table 2 and Figure A.3), we find in our  $\hat{\gamma}$  values that our Bayesian MF Weibull model generally recovers each true  $\gamma$  value quite well. That being said, the RMSE, CP, and MCSE values reported in Table 2 nevertheless suggest that the Bayesian MF Weibull model’s  $\hat{\gamma}$ ’s exhibit slightly higher bias, worse coverage, and lower efficiency than was the case for the Bayesian MF Weibull’s  $\hat{\beta}$ ’s in Experiment 2.

We next turn to MC Experiments 3–4, which assess the performance of our maximum likelihood estimated (BFGS) Weibull and MF Weibull models in circumstances where one’s outcome variable follows a Weibull survival process (Experiment 3) or a MF Weibull survival process (Experiment 4). We report these full MC results in the Supplementary Appendix, and summarize the key insights here. First and foremost, Experiments 3–4 yield similar conclusions to those obtained in Experiments 1–2. When one’s d.g.p. is Weibull (Experiment 3), the BFGS Weibull and BFGS MF Weibull models perform comparably, with no noticeable differences in bias, coverage, or efficiency across these two estimators. However, when the d.g.p. is instead MF Weibull (Experiment 4), the BFGS MF Weibull exhibits consistently lower bias, superior coverage, and higher efficiency than the BFGS Weibull model, with the BFGS MF Weibull’s RMSEs generally being 5–10 times smaller than those of the BFGS Weibull. Hence, the risks to inference of (mis)applying the MF Weibull in the absence of MFs are fairly low, whereas the inferential risks of (mis)applying a Weibull to MF survival data are substantial.

We can also compare the Bayesian MF Weibull results obtained in Experiment 2 to those of the BFGS MF Weibull in Experiment 4. Here we observe that the  $\hat{\beta}$ ’s from each MF model are comparable across Experiments 2 and 4, as are the corresponding RMSEs. However, when one’s outcome variable is MF Weibull, the Bayesian MF Weibull model’s  $\hat{\gamma}$ ’s generally exhibit lower bias, and higher efficiency, than those of the BFGS MF Weibull. Thus, we can conclude that the Bayesian MF Weibull is superior to the BFGS MF Weibull model in accuracy and efficiency when the d.g.p. is MF Weibull. This suggests that researchers should generally favor the Bayesian MF Weibull model over the BFGS MF Weibull model for applied research.

Experiments 5–8 simulate either an exponentially distributed outcome variable (Experiments 5 and 7) or a MF exponential outcome variable (Experiments 6 and 8). These experiments are fully presented in the Supplementary Appendix, and reevaluate our Bayesian or BFGS (MF) Weibull models alongside Bayesian or BFGS (MF) exponential survival models. Across MC Experiments 5–8, we again find that the (Bayesian and BFGS) MF survival models perform comparably to appropriate non-MF survival estimators when the true d.g.p. exhibits no MFs. When the d.g.p. is instead MF exponential, we determine that (i) the (Bayesian and BFGS) MF survival models again substantially outperform all non-MF survival models in bias, CPs, and efficiency and (ii) the Bayesian MF survival models generally remain preferable to the BFGS MF survival models in these contexts. Furthermore, we find in each relevant comparison that the MF *Weibull* models exhibit comparable, and at times superior, performance to the MF *exponential* models. This suggests that the Weibull MF model should be preferred over the MF exponential estimator in applied research, given the former’s added flexibility in situations where one’s hazard rate is non-constant.

Whereas Experiments 2, 4, 6, and 8 employ a MF rate of 5%, Experiments 9–11 increase this MF rate above 5%. These latter experiments—which we present in the Supplementary Appendix—increasingly favor the Bayesian MF Weibull model over the Bayesian Weibull model as one’s MF rate extends beyond 5%. To illustrate this, we average the  $N = 1,000$ ,  $N = 1,500$ , and  $N = 2,000$  RMSE results that we obtain from Experiment 1 ( $\alpha = 0$ ), Experiment 2 ( $\alpha = 5\%$ ), Experiment 9



**Figure 1.** Comparison of survival stage RMSEs under increasing MF rates.

( $\alpha = 8\%$ ), Experiment 10 ( $\alpha = 10\%$ ), and Experiment 11 ( $\alpha = 15\%$ ). We then plot these averaged RMSE values (and their standard deviations) separately for our Bayesian Weibull and Bayesian MF Weibull models in Figures 1a ( $\hat{\beta}_0$ ) and 1b ( $\hat{\beta}_1$ ). Herein, both models exhibit comparable RMSEs for  $\hat{\beta}_0$  and  $\hat{\beta}_1$  when the d.g.p. is non-MF Weibull. However, as one's MF rate increases, we find that the Bayesian MF Weibull model's  $\hat{\beta}$  RMSEs remain effectively flat, whereas those the Bayesian Weibull dramatically increase. Notably, the Bayesian MF Weibull already exhibits RMSEs that are over 5 times smaller than those of Bayesian Weibull when  $\alpha = 5\%$ , and these Bayesian MF Weibull RMSEs then become up to 28 times smaller than those of the Bayesian Weibull when  $\alpha$  is increased to 15%. Hence, Experiments 9–11 further underscore the preferability of the Bayesian MF Weibull in situations of modest-to-moderate MFs.

The Supplementary Appendix also evaluates the sensitivity of our MC findings to two additional scenarios. First, Experiments 12–13 reevaluate our primary MC findings when our Bayesian (MF) Weibull models are specified with a multivariate Cauchy prior as opposed to our favored multivariate normal prior. This choice does not affect the substantive insights discussed above, although the multivariate Cauchy prior does yield slightly higher bias in the Bayesian MF Weibull's  $\gamma$  estimates, relative to the values reported in Table 2. Second, Experiments 14–15 in the Supplementary Appendix examine the performance of the Bayesian (MF) Weibull models when applied to a (MF) log-logistic d.g.p., and compare these models to a Bayesian Cox proportional hazards (PH) estimator that makes no assumptions about the shape of the baseline hazard function. When an outcome's d.g.p. is log-logistic, we find that the Bayesian Cox PH model outperforms the Bayesian (MF) Weibull models. However, when an outcome's d.g.p. is log-logistic with a 5% MF rate, our Bayesian MF Weibull model noticeably outperforms the Bayesian Cox PH model in both bias and coverage. Thus, in instances where one encounters a non-Weibull distributed outcome variable that exhibits a modest MF rate, the Bayesian MF Weibull model can often remain a superior choice over non-MF, semiparametric alternatives.

#### 4 Empirical Applications

We estimate our Bayesian MF Weibull model on survival data used in two published studies in Political Science that employ standard Weibull models. The second application—fully presented in the Supplementary Appendix—is about the survival of democratic regimes. However, for our first application presented here, we consider a survival dataset measuring the duration of civil conflicts obtained from a study published by Buhaug, Gates and Lujala (hereafter Buhaug *et al.*)

in 2009. Their paper theoretically posits that geographic covariates such as logged distance from the conflict center to the capital city (*Distance to capital (ln)*) and civil conflicts in border regions (*Conflict at border*) decrease the hazard of civil war termination or equivalently lead to longer civil wars, while higher *Rebel fighting capacity* increases the hazard of civil war failure. More importantly, following Collier, Hoeffler, and Söderbom (2004) and Fearon and Laitin's (2003) theoretical claim, Buhaug *et al.* (2009) also assess whether countries with higher *GDP per capita at onset (ln)* of civil wars are likely to be associated with a higher hazard of civil war termination (i.e., shorter civil conflicts). Next, they test—as suggested by Balcells and Kalyvas (2014)—whether the dummy for *post-Cold War* years is associated with a higher hazard of civil war failure. Finally, following Cunningham (2006) Buhaug *et al.* (2009: 551–554) test whether *higher democracy score at (the) onset* of civil wars are associated with longer civil conflicts.

To statistically assess these theoretical predictions, Buhaug *et al.* (2009) use country-level survival data measuring the duration of civil conflicts (1946–2003) in days as the outcome variable, which is labeled as *civil war duration*. These data are obtained from the Uppsala/PRIO Armed Conflict Dataset (ACD). Building on extant civil war duration analyses (e.g., Balch-Lindsay and Enterline 2000, Thyne 2012), civil conflict is coded as “terminated” by Buhaug *et al.* (2009: 556) when the number of battle deaths *falls and stays* below—as per the UCDP/PRIO ACD criterion—the threshold of 25 for at least 24 months. Buhaug *et al.* (2009) estimate a standard parametric maximum likelihood estimation (MLE) Weibull model in which they include the following covariates listed earlier that influence civil war duration: *distance to capital (ln)*, *conflict at border*, a binary measure of *rebel (group) fighting capacity*, *GDP capita at onset (ln)* of civil conflicts, the *post-Cold War* years dummy, a measure of *democracy score at onset* of civil conflicts, and a *Border × distance (ln)* control. In a fully specified MLE Weibull model, Buhaug *et al.* (2009: 563) find support for their predictions that *conflict at border*, *distance to capital (ln)* and *democracy score at onset* statistically have a negative (positive) and reliable effect on the hazard of civil conflict failure (civil conflict duration). They also report that the statistical association between *GDP capita at onset (ln)* and the hazard of civil conflict termination is positive, but unreliable. They, however, find robust support for the claims that *post-Cold War* and *rebel fighting capacity* have a statistically positive and reliable effect on the hazard of civil conflict failure.

Although now standard practice in the civil conflict literature (Thyne 2012; Themnér and Wallensteen 2014), Buhaug *et al.*'s use of an annual 25 battle-deaths threshold over a 24-month period as a criterion to code conflict termination can lead to the inclusion of MF cases in the data. First, the use of 24-month spells to identify conflict termination is likely to be *conservative* for lower-intensity conflicts in remote or poor information environments, or in situations where some groups or officials do not recognize the conflict as having ended. Such cases where the date of civil conflict termination is ambiguous are unlikely to capture the “true” end date. Take, for instance, the Second Congo War, which officially ended in 2003. Is the correct end date July 2003 during which a provincial government assumed power? Or is October 2008, the date recorded for termination of the Second Congo War in the UCDP Conflict Termination Dataset (Kreutz 2010), more accurate although other key sources suggest that Congo's civil war has not ended (Larmer, Laudati, and Clark 2013)?

A second issue is that different sources may record distinct dates for civil conflict termination even though they use the same battle-death numbers threshold criterion to code the “end” of civil conflict because of subjectivity involved in accurately identifying the number of battle deaths. For instance, the UCDP Conflict Termination Dataset (Kreutz 2010) used by Buhaug *et al.* (2009) denotes a civil conflict in the state of Nagaland in India as beginning in 1992 and experiencing termination in 1997, the first year during which the number of battle deaths fell below 25. Yet other sources that use the same UCDP battle-death threshold criterion emphasize that civil conflict between the Indian Government and Nagaland's rebels during the 1990s did not “end” in 1997 but

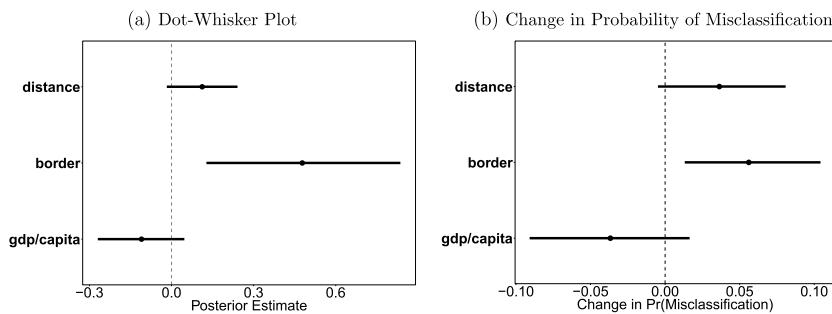
rather persisted into 2004 or beyond (Shimray 2001). *Table A.13* in the Supplementary Appendix also identifies many additional terminated civil conflict cases in the Buhaug *et al.* data—e.g., civil wars in other parts of India, Myanmar, the Democratic Republic of Congo, and Thailand—that persisted beyond their recorded failure time. Hence, without perfect information about civil war termination dates and the number of civil conflict battle deaths, it is plausible that civil war duration datasets including Buhaug *et al.*’s data are contaminated with MF cases that have persisted beyond their observed-failure point.

We thus replicate a key specification from Buhaug *et al.* (2009; Table 1, Column 5) by separately estimating and comparing the results from the following models: (i) our Bayesian MF Weibull model that (unlike the standard Weibull models) statistically accounts for MF cases in Buhaug *et al.*’s civil conflict duration data, (ii) a standard MLE Weibull model and (iii) a standard Bayesian Weibull model. We present the results from each of these models graphically below and in the Supplementary Appendix (to save space). Note that these models focus on the Buhaug *et al.* (2009) specification which evaluates the following variables’ effect on *civil conflict duration*: *GDP capita at onset (ln)*, *post-Cold War years*, *democracy score at onset*, *rebel fighting capacity*, *distance to capital (ln)*, *conflict at border*, and *Border × distance (ln)*. Specifically, we start our assessment of this Buhaug *et al.* (2009) specification by estimating a standard Weibull hazard model first via MLE and then via Bayesian MCMC. Additionally, we estimated four different Bayesian MF Weibull model specifications. To this end, recall that unlike the standard Weibull model, the Bayesian MF Weibull model estimates the effect of both (i) a series of **X** covariates on *civil conflict duration*, and (ii) a set of **Z** covariates on the probability of failure misclassification (denoted as  $\alpha$ ).

Hence, for the Bayesian MF models, we first report a baseline Bayesian MF Weibull specification. The survival stage in this baseline MF model of *civil conflict duration* includes the same variables used in the Buhaug *et al.* (2009) study, while the misclassification failure probability stage (hereafter “misclassification stage”) includes just the intercept. The survival stage of the second Bayesian MF Weibull specification also incorporates the same variables used by Buhaug *et al.* (2009), but adds some theoretically identified covariates to the MF model’s misclassification stage. Here, we first include *GDP capita at onset (ln)* since conflict-afflicted countries with higher levels of economic development may have greater media coverage of civil conflicts (Collier 2003, Puddephatt 2006). This improves the accuracy of information about civil conflict termination dates as per the UCDP battle-deaths criterion, which reduces the probability of misclassification failure. Next, we include *distance to capital (ln)* as information about battle-related fatalities (needed to code civil war termination) is often inaccurate in civil conflicts fought in remote areas far away from the capital city (Puddephatt 2006). This misclassification stage covariate is thus likely to be positive. We also incorporate *conflict at border* in the misclassification stage and expect it to be positive as governments in civil war-affected countries often provide inaccurate information about battle-related deaths in civil wars in their state’s border regions owing to security concerns (Buhaug and Gates 2002, Lischer 2015).

The survival stage of the third Bayesian MF Weibull specification also repeats the survival stage used by Buhaug *et al.* (2009; Table 1, Column 5), while this specification’s misclassification stage includes the three covariates discussed above and the following two variables: *rebel fighting capacity* and *democracy at onset*. Finally, the fourth specification includes all covariates from the survival stage in the Buhaug *et al.* study for both the Bayesian MF Weibull model’s survival stage and misclassification stage. The Bayesian MF Weibull (and Bayesian Weibull) models are each estimated using the multivariate normal prior and our slice-sampling (MCMC) algorithm for which we specify the hyperparameters as:  $a = 1$ ,  $b = 1$ ,  $S_\beta = I_{p1}$ ,  $S_\gamma = I_{p2}$ ,  $v_\beta = p1$  and  $v_\gamma = p2$ .<sup>6</sup> The results remain robust when these models are estimated using the multivariate Cauchy

<sup>6</sup> The Bayesian MF Weibull specification results are based on a set of 50,000 iterations (10,000 iterations in one case) after 3,000 burn-in scans.



**Figure 2.** Gamma ( $\gamma$ ) covariate results.

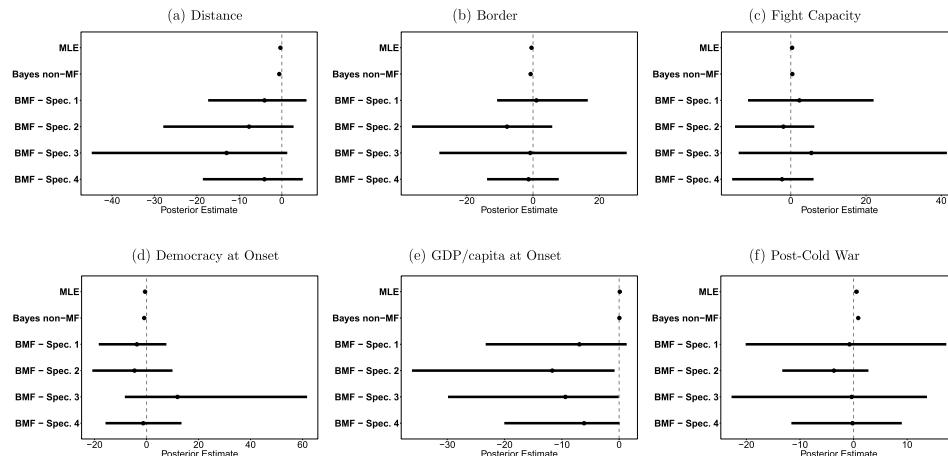
prior (see Section IV, Supplementary Appendix). We first discuss the Bayesian MF Weibull model's misclassification stage and then the model's survival stage results. The misclassification stage results are presented via the following illustrations derived from the models results: dot-whisker plots including Figure 2a and Figures A.25(a)–(b) in the Supplementary Appendix that illustrate each misclassification stage covariate's posterior mean estimate with its 95% Bayesian Credible Intervals (hereafter "BCI"), the first difference in misclassification probabilities from the Bayesian MF model's misclassification stage ( $\mathbf{Z}$ ) covariates (Figure 2b), and the posterior probability of some misclassification stage covariates derived from their respective posterior sample estimates.

The misclassification stage covariates dot-whisker plots mentioned above show that *conflict at border* is positive in the misclassification stage of each Bayesian MF Weibull specification and this estimate's 95% BCIs exclude zero. The first difference in misclassification probabilities derived from the second Bayesian MF Weibull specification's misclassification stage covariates reveals that increasing the *conflict at border* dummy from 0 to 1 (here and below, while other covariates are held at their means or modes) increases the probability of a misclassified war failure by approximately 5.61% and the 95% BCI of this effect excludes zero (see 2b). Moreover, the posterior probability that the hypothesized effect of conflict at border on the likelihood of MF is positive is 0.996. Hence, as predicted theoretically, it is reliable to infer that civil conflicts that occur in the border regions of war-torn countries are more likely to be misclassified as having been terminated when they (possibly) had not. As anticipated, the posterior mean estimate (Figure 2a), posterior probability and the substantive effect of *distance to capital (ln)* in the misclassification stage (Figure 2b) show that civil conflicts fought in geographically remote areas are indeed more likely to be misclassified as having failed when they had not. But the 95% BCI of this covariate's posterior mean estimate illustrated in Figure 2a and Figure A.25(a) includes zero. Thus, the aforementioned empirical relationship is not reliable.

The posterior mean estimate (Figure 2a and Figures A.25(a)–(b)), posterior probability and the substantive effect of *GDP capita at onset (ln)* in Figure 2b show that the relationship between this covariate and the likelihood of misclassification failure is negative (as hypothesized) in the Bayesian MF Weibull specification's misclassification stage. But the 95% BCI of this covariate's mean estimate includes zero in most—but excludes zero in one of the—misclassification stage specifications. While this result for *GDP capita at onset (ln)* supports our claim that civil conflicts in more economically developed countries are less likely to be misclassified as having been terminated when they had not, it also shows that this empirical association is less reliable.

We next discuss the survival stage covariate estimates in the (MF) Weibull models that focus on the hazard of civil conflict termination. To this end, first consider the top two specifications in each of the dot-whisker plots<sup>7</sup> in Figures 3a–3f that illustrate the survival stage (i) coefficient estimates

<sup>7</sup> The dot-whisker plots for the constant and the interaction term variable are illustrated in Figures A.26a–A.26b of the Supplementary Appendix.



**Figure 3.** Dot-whisker plots for beta covariates in MLE and Bayesian (MF) Weibull models.

of the covariates from the standard MLE Weibull model and their respective 95% confidence intervals and (ii) posterior mean estimates of the same covariates from the Bayesian non-MF Weibull model and their respective 95% BCI. Figures 3a, 3b and 3d show that the influence of *distance to capital (ln)*, *conflict at border* and *democracy score at onset* on the hazard of civil conflict termination are each negative and highly reliable in the standard MLE and Bayesian Weibull models, which is exactly what Buhaug *et al.* (2009) finds for these covariates. The positive effect of *rebel fighting capacity* on the hazard of conflict failure is also reliable in the MLE Weibull model, as shown by Buhaug *et al.* (2009: 561). *GDP capita at onset (ln)* is positive in both the standard MLE and non-MF Bayesian Weibull models (Figure 3e), which mirrors Buhaug *et al.*'s (2009: 563) result for this covariate. The positive influence of *GDP capita at onset (ln)* is statistically unreliable in the MLE Weibull specification as (also) found by Buhaug *et al.* (2009), but statistically reliable in the Bayesian non-MF Weibull model. Figure 3f shows that the positive association between *post-Cold War* and the hazard of civil conflict termination is statistically reliable in the MLE and Bayesian non-MF Weibull models, which is what Buhaug *et al.* (2009) and Balcells and Kalyvas (2014) find.

However, the Bayesian MF Weibull's estimates differ substantially from those in the standard MLE (and Bayesian) Weibull models that Buhaug *et al.* (2009) report.<sup>8</sup> To see this, we focus on the: (i) bottom four specifications in Figures 3a–3f that illustrate the survival stage covariates' posterior mean estimates and their respective 95% BCI from the four Bayesian MF Weibull specifications delineated earlier and (ii) posterior probability of the key Bayesian MF Weibull model's survival stage covariates reported that are derived from their respective posterior sample estimates<sup>9</sup>. We also assess the hazard ratio plots of the Bayesian MF Weibull model's key survival stage covariates in Figures A.27(a)–(b) in the Supplementary Appendix.

The survival stage posterior mean estimates (Figures 3a–3b) of both *distance to capital (ln)* and *conflict at border* show that each of these two covariates are almost always negatively associated with the hazard of civil conflict failure in the MF Weibull models, although this association is unreliable since the 95% BCIs of the variables include zero. This result is distinct from Buhaug *et al.* (2009) who find that the negative association between each of these two covariates and the hazard of conflict failure is highly robust in their MLE Weibull model. This suggests that once we

<sup>8</sup> Beyond the differences in effect size and/or direction (discussed below), note that our Bayesian MF Weibull survival estimates also are less reliable than our standard Weibull estimates, given that the former survival stage estimates are now conditioned, during estimation, on a set of MF estimates (and corresponding MF probabilities) that exhibit their own levels of imprecision.

<sup>9</sup> The posterior probability is calculated by dividing the total number of posterior estimates that confirm the hypothesized effect of the covariate by the total number of posterior estimates. A higher posterior probability confirms the hypothesized relationship more reliably.

statistically account for MF cases in the Bayesian MF Weibull model, Buhaug *et al.*'s hypothesized negative relationship between the two covariates—*distance to capital (ln)* and *conflict at border*—and the hazard of civil war failure is weak.

Next, the survival stage posterior mean estimate of log of *GDP capita at onset* in the baseline Bayesian MF Weibull specification (where the misclassification stage only includes the intercept) is negative (see Figure 3e). The survival stage estimate of *GDP capita at onset (ln)* remains negative and its 95% BCI excludes zero in the Bayesian MF Weibull specifications in which the misclassification stages include other covariates (Figure 3e). Further, based on (for example) the third Bayesian MF Weibull specification's survival stage results, the posterior probability that the log of per capita income's (onset) effect on the hazard of civil conflict termination is positive, as predicted by Buhaug *et al.* (2009), is merely 0.022. Hence, the Bayesian MF Weibull model's survival stage results indicate that the log of per capita income at the onset of civil wars has a reliably negative influence on the hazard of civil war failure, which is exactly the opposite of what Buhaug *et al.* (2009) find. This suggests that a possible *prolonging* effect of GDP per capita on civil war duration may have gone unnoticed in many past analyses,<sup>10</sup> which failed to take into account MF cases and these MF cases' potential association with low GDP per capita.

Additionally, consider Figure A.27(a) in the Supplementary Appendix. This figure illustrates hazard ratio plots derived from the estimate of *GDP capita at onset (ln)* in the third Bayesian MF Weibull specification that includes theoretically identified covariates and additional controls in the misclassification stage, standard MLE and Bayesian Weibull specifications. Figure A.27(a) reveals that increasing *GDP capita at onset (ln)* from 1 SD below to 1 SD above its mean while holding the other survival stage covariates at their mean or mode increases the hazard of civil war termination in the standard MLE and Bayesian Weibull models; this effect is, however, unreliable in the MLE Weibull model. In sharp contrast, the figure shows that increasing *GDP capita at onset (ln)* from 1 SD below to 1 SD above its mean reliably decreases the hazard of civil conflict termination by 32.2% in the Bayesian MF Weibull model. Hence, reasonable increases in per capita income at the outbreak of civil wars increase the hazard of civil conflict failure in the standard Weibull models. But the same changes in *GDP capita at onset (ln)* in the Bayesian MF Weibull specification lead to a substantial and reliable decrease in the hazard of civil war termination after MFs are accounted for.

We now assess another variable, the survival stage estimate of *post-Cold War* years. As predicted by Balcells and Kalyvas (2014) and Buhaug *et al.* (2009), the hazard ratio plot in Figure A.27(b) in the Supplementary Appendix shows that increasing the *post-Cold War* dummy from 0 to 1 while holding other survival stage covariates at their mean or mode increases the hazard of civil war termination in the standard MLE and Bayesian Weibull models and this effect is highly reliable. But in contrast to the standard Weibull model's results for this variable, the posterior mean estimate of *post-Cold War* is—as illustrated in Figure 3f—negative in the survival stage<sup>11</sup> in the Bayesian MF Weibull specifications and the 95% BCI of this covariate includes zero in all of these specifications.<sup>12</sup> Moreover, the effect of changing the *post-Cold War* dummy from 0 to 1 on the hazard of civil conflict termination (while holding the other survival stage covariates at their mean or mode) is highly unreliable as well. Hence, unlike previous studies, the association between *post-Cold War* and the hazard of civil war termination tends to be negative but this

<sup>10</sup> Exceptions include Balcells and Kalyvas (2014), who find that GDP per capita increases civil war duration (although this result is unreliable in some of their specifications), as well as Brandt *et al.* (2008), who find that GDP per capita increases civil war duration for civil wars ending in government victory specifically.

<sup>11</sup> Interestingly, we learn from the third Bayesian MF Weibull specification's survival stage (for example) posterior probability that the hypothesized effect of *post-Cold War* on the hazard of civil war failure is negative is almost 0.98, which is distinct from the Buhaug *et al.* (2009) finding for this covariate.

<sup>12</sup> We also learn from the Bayesian MF Weibull specification's survival stage results and posterior probability that *post-Cold War*'s effect on the hazard of civil war termination is positive, as suggested by Balcells and Kalyvas (2014).

result is not consistently reliable. Last, we find similar contradictory results for additional survival stage covariates from the Buhaug *et al.* (2009) study, including (for example) *democracy score at onset*, suggesting that past findings in these regards may have at least been partly attributable to misclassification in civil conflict failures, and to associations between failure misclassification and the variables reviewed here.

A battery of specification robustness tests presented in the Supplementary Appendix shows that the posterior mean estimates of all the key survival stage covariates in the Bayesian MF Weibull specifications such as per capita income, post-Cold War years, and democracy score remain robust in not just additional specifications that include alternative controls in the MF Weibull model's misclassification stage but also specifications estimated using the Cauchy prior (see Section IV of Supplementary Appendix). Next, as described in the Supplementary Appendix, convergence diagnostic checks including autocorrelation plots, the Geweke (1992) convergence diagnostic test and the Heidelberger and Welch (1983) test of stationarity show that in the Bayesian MF Weibull specifications of interest, (i) all the misclassification stage parameter estimates have converged properly and (ii) all the survival stage parameter estimates (including *GDP capita at onset (ln)* and *post-Cold War*) barring one have also converged properly (Figures A.28–A.29 and Tables A.14–A.15, Supplementary Appendix). Hence, altogether, comparisons of the standard Weibull and Bayesian MF Weibull models suggest that the effects of several widely used predictors of civil war duration are sensitive to the potential misclassification of civil war cases as having failed when in fact they persisted. After statistically accounting for MFs within one widely used dataset of civil war duration, we find that theoretical interpretations of some correlates of civil war duration reverse in sign whereas others change in magnitude and/or become less reliable. Thus, more attention should be paid to the “underreporting” of civil war termination dates in empirical conflict research. The MF models proposed above allow researchers to account for such underreporting of failure dates and assess when failure cases in survival datasets are more likely to be misclassified, which is substantively appealing and empirically useful.

To further validate our model, we compare the predicted probability of MF derived from our Bayesian MF Weibull's misclassification stage to the UCDP/PRIO's precision coding records of the civil war end dates for the conflict years in Buhaug *et al.* (2009). To this end, first note that UCDP/PRIO codes the degree of precision of civil war end dates in the ACD used by Buhaug *et al.* (2009) on an ordinal 1 (all elements of the end date—day, month, year—are accurate) to 5 (at least two date elements are unknown and thus imprecise) scale. We use this precision coding to operationalize two binary variables. The first (*Imprecision\_1*) is coded as 1 for civil war end dates in the Buhaug *et al.* (2009) data that are equal to 5 in the precision scale mentioned above and 0 otherwise. The second binary variable (*Imprecision\_2*) captures civil war end dates that are greater than or equal to 3 in the 1–5 precision coding scale and is coded as 0 otherwise. Therefore, *Imprecision\_1* denotes rebel-government civil war end dates that are highly imprecise as per UCDP/PRIO's precision coding criteria, whereas *Imprecision\_2* denotes rebel-government civil war end dates that are highly or moderately imprecise as per UCDP/PRIO's precision coding criteria.

After operationalizing these two binary variables, we employ the misclassification stage estimates from the full Bayesian MF Weibull specification to derive the predicted probability of a civil war's MF for each conflict year in Buhaug *et al.* (2009). We then evaluate the accuracy by which these MF probabilities classify the two binary variables described above, using areas under the receiver operating characteristic curve (AUCs) and F1 scores. In undertaking these comparisons, we find that we are able to classify *Imprecision\_1* with an AUC of 0.62 and an F1 Score of 0.52, whereas we are able to predict *Imprecision\_2* with an AUC of 0.63 and an F1 Score of 0.55. Thus, our misclassification stage does only a modest job in predicting the two imprecision indicators, but does nevertheless offer some predictive leverage. Given that our

misclassification stage is only specified with the variables used by Buhaug *et al.*, and given that our predictions do not come from a model that was itself directly trained on the binary imprecision outcomes evaluated here, these current levels of accuracy are perhaps more impressive than they would initially seem. Furthermore, the fact that our classification statistics each improve when we include moderate imprecision cases in our binary outcome measure (i.e., for *Imprecision\_2*) suggests that our misclassification stage's accuracy in classifying imprecision improves as the “1's” on this imprecision indicator more fully encompass the relevant imprecision cases in Buhaug *et al.* (2009)—thereby further helping to validate our model and application.

## 5 Conclusion

Event failures in Political Science survival datasets are often imperfectly recorded according to crude cutoff criteria or related misreporting processes. Imperfectly recorded event failures ensure that some non-censored observations actually persist beyond their *recorded failure* in a survival dataset. When this arises, conventional survival models can yield biased estimates. To address this problem, we build on recent work on split population survival models and develop a new “misclassified failure” (MF) split population survival model that explicitly models the probability of MF (vs. right-censored) events. In doing so, our model accounts for imperfect detection in failure events within one's evaluations of covariate effects on survival (i.e., duration) processes. As a result, the MF split population survival model provides more accurate estimates of the parameter effects when observed event failures include cases that in actuality “live on” past their observed-failure point.

We define this model's conditional posterior distribution and present a slice-sampling estimation algorithm that allows researchers to conduct Bayesian inference on our model. We also provide a dedicated R package for estimating this Bayesian MF survival model as a complement to this paper. Results from extensive MC experiments and two empirical applications reveal that when some recorded event failures in survival data have survived past their observed-failure points, our Bayesian MF model yields estimates that are superior in accuracy and coverage compared to estimates from regular survival models. Our MF duration model also provides researchers with an opportunity to include variables in not only the model's survival stage but also within a stage that models the probability of a MF. This allows one to identify the conditions that affect whether a duration case is either more or less likely to be misclassified as having terminated; potentially providing substantive insights into this secondary process. For some applications, these insights will help to inform researchers of problematic coding and data collection decisions with respect to event failures. In other cases, these insights and the substantive effects derived from our model may reveal the theoretical mechanisms that cause political actors to overstate failure in some cases but not others.

Notwithstanding these benefits, the model presented here can be extended in three main directions. First, our statistical framework can be potentially used to develop the semiparametric Cox PH MF model. Although scholars in Comparative Politics and International Relations commonly use parametric survival models such as the Weibull model considered above, Political Scientists also frequently use Cox PH models. It is plausible that our parametric MF duration model could be extended to the Cox PH context. Second, we focused on two empirical applications in our paper: civil war duration and the survival of democratic regimes. Yet we mentioned earlier that other survival datasets analyzed by scholars (e.g., Cress, McPherson, and Rotolo 1997; Cioffi-Revilla and Landman 1999; Box-Steffensmeier, Radcliffe, and Bartels 2005) could also include imperfectly recorded event failures that have survived past their observed-failure points. It may thus be worthwhile to apply our parametric MF duration models to statistically assess additional duration outcomes. Finally, we also note that left-censored survival data is widespread in Comparative

Politics and International Relations (Carter and Signorino 2013). As such, our MF survival model could also be extended to address this issue.

## Supplementary material

For supplementary material accompanying this paper, please visit

<https://doi.org/10.1017/pan.2019.6>.

## References

Bagozzi, B., M. Joo, B. Kim, and B. Mukherjee. 2019. "Replication Data for: A Bayesian Split Population Survival Model for Duration Data With Misclassified Failure Events." <https://doi.org/10.7910/DVN/ILZHUR>, Harvard Dataverse, V1.

Balcells, L., and S. N. Kalyvas. 2014. "Does Warfare Matter? Severity, Duration, and Outcomes of Civil Wars." *Journal of Conflict Resolution* 58(8):1390–1418.

Balch-Lindsay, D., and A. J. Enterline. 2000. "Killing Time: The World Politics of Civil War Duration, 1820–1992." *International Studies Quarterly* 44(4):615–642.

Beger, A., C. L. Dorff, and M. D. Ward. 2014. "Ensemble Forecasting of Irregular Leadership Change." *Research & Politics*, 1–14.

Beger, A., C. L. Dorff, and M. D. Ward. 2016. "Irregular Leadership Changes in 2014: Forecasts Using Ensemble, Split-population Duration Models." *International Journal of Forecasting* 32(1):98–111.

Beger, A., D. W. Hill, N. W. Metternich, S. Minhas, and M. D. Ward. 2017. "Splitting It Up: The spduration Split-Population Duration Regression Package for Time-Varying Covariates." *The R Journal* 9(2):474–486.

Box-Steffensmeier, J. M., and C. Zorn. 1999. "Modeling Heterogeneity in Duration Models." Paper presented at the 1999 Summer Meeting of The Political Methodology Society.

Box-Steffensmeier, J. M., P. M. Radcliffe, and B. L. Bartels. 2005. "The Incidence and Timing of PAC Contributions to Incumbent U.S. House Members, 1993–94." *Legislative Studies Quarterly* 30(4):549–579.

Box-Steffensmeier, J. M., S. De Boef, and K. Joyce. 2007. "Event Dependent and Heterogeneity in Duration Models: the Conditional Frailty Model." *Political Analysis* 15(3):237–256.

Brandt, P. T., T. D. Mason, M. Gurses, N. Petrovsky, and D. Radin. 2008. "When and How the Fighting Stops: Explaining the Duration and Outcome of Civil Wars." *Defence and Peace Economics* 19(6):415–434.

Buhaug, H., and S. Gates. 2002. "The Geography of Civil War." *Journal of Peace Research* 39(4):417–433.

Buhaug, H., S. Gates, and P. Lujala. 2009. "Geography, Rebel Capability, and the Duration of Civil Conflict." *Journal of Conflict Resolution* 53(4):544–569.

Carlin, B. P., and T. A. Louis. 2000. *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd edition. New York: Chapman and Hall/CRC.

Carter, D. B., and C. S. Signorino. 2013. "Good Times, Bad Times: Left Censoring in Grouped Binary Duration Data." Presented at the Meeting of the International Studies Association, 3–6 April, 2013.

Cioffi-Revilla, C., and D. Lai. 1995. "War and Politics in Ancient China, 2700 B.C. to 722 B.C.: Measurement and Comparative Analysis." *Journal of Conflict Resolution* 39(3):467–494.

Cioffi-Revilla, C., and T. Landman. 1999. "Evolution of Maya Polities in the Ancient Mesoamerican System." *International Studies Quarterly* 43(4):559–598.

Collier, P. 2003. "The Market for Civil War." *Foreign Policy* (136):38–45.

Collier, P., A. Hoeffler, and M. Söderbom. "On the Duration of Civil War." *Journal of Peace Research* 41(3):253–273.

Cress, D. M., J. M. McPherson, and T. Rotolo. 1997. "Competition and Commitment in Voluntary Memberships: The Paradox of Persistence and Participation." *Sociological Perspectives* 40(1):61–79.

Cunningham, D. E. 2006. "Veto Players and Civil War Duration." *American Journal of Political Science* 50(4):875–892.

Fearon, J. D., and D. Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97(1):75–90.

Geweke, J. 1992. "Evaluating the Accuracy of Sampling Based Approaches to the Calculation of Posterior Moments." In *Bayesian Statistics*, vol. 4, edited by J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 169–193. New York: Oxford University Press.

Heidelberger, P., and P. D. Welch. 1983. "Simulation Run Length Control in Presence of an Initial Transient." *Operations Research* 31:1109–1144.

Jin, S., and F. J. Boehmke. 2017. "Proper Specification of Nonproportional Hazards Corrections in Duration Models." *Political Analysis* 25(1):138–144.

Keele, L. 2010. "Proportionally Difficult: Testing for Nonproportional Hazards in Cox Models." *Political Analysis* 18(2):189–205.

Kreutz, J. 2010. "How and When Armed Conflicts End: Introducing the UCDP Conflict Termination Dataset." *Journal of Peace Research* 47(2):243–250.

Larmer, M., A. Laudati, and J. F. Clark. 2013. "Neither War Nor Peace in the Democratic Republic of Congo (DRC): Profiting and Coping Amid Violence and Disorder." *Review of African Political Economy* 40:1–12.

Licht, A. A. 2011. "Change Comes with Time: Substantive Interpretation of Nonproportional Hazards in Event History Analysis." *Political Analysis* 19(2):227–243.

Lischer, S. K. 2015. *Dangerous Sanctuaries: Refugee Camps, Civil War, and the Dilemmas of Humanitarian Aid*. Ithaca, NY: Cornell University Press.

Metzger, S. K., and B. T. Jones. 2016. "Surviving Phases: Introducing Multistate Survival Models." *Political Analysis* 24(4):457–477.

Neal, R. M. 2003. "Slice Sampling." *The Annals of Statistics* 31(3):705–741.

Puddephatt, A. 2006. "Conflict and the Role of the Media." Copenhagen: International Media Support Report.

Ruhe, C. 2018. "Quantifying Change Over Time: Interpreting Time-varying Effects In Duration Analyses." *Political Analysis* 26(1):90–111.

Shimray, U. A. 2001. "Ethnicity and Socio-political Assertion: the Manipur Experience." *Economic and Political Weekly* 36(39):3674–3677.

Svolik, M. W. 2008. "Authoritarian Reversals and Democratic Consolidation." *American Political Science Review* 102(2):153–168.

Themnér, L., and P. Wallensteen. 2014. "Armed conflicts, 1946–2013." *Journal of Peace Research* 51(4):541–554.

Thyne, C. L. 2012. "Information, Commitment, and Intra-War Bargaining: The Effect of Governmental Constraints on Civil War Duration." *International Studies Quarterly* 56(2):307–321.

Young, J. K., and L. Dugan. 2014. "Survival of the Fittest: Why Terrorist Groups Endure." *Perspectives on Terrorism* 8(2):2–23.