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

March 7, 2018

Minnie Joo,* Benjamin E. Bagozzi,† Bomin Kim,‡ Bumba Mukherjee§

Abstract

Researchers often study outcome variables that correspond to 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. Within studies of civil war duration, for example, event failures are imperfectly identified according to a crude cutoff criteria, ensuring that some civil wars that are coded as terminated (i.e., as *non-right censored*) persist beyond their *recorded failure*. Inaccurately recorded event failures of this sort 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 terminated at that point in time is problematic as there is a non-zero probability that the observation persisted past that point. Moreover, if heterogeneity exists among these imperfect codings of event failures, then survival models will yield biased estimates of parameter effects. To address this problem we develop a new split population survival estimator that explicitly models the misclassification probability of failure (vs. right censored) events. After deriving this model, and an associated R package, we use Bayesian estimation via a slice sampling algorithm to evaluate its performance within both (i) simulated data and (ii) several published political science applications. We find that our proposed "misclassified failure" survival model allows researchers to accurately account for the process of "inflation" in failure-events that is described above.

*Ph.D. Candidate, Dept. of Political Science, Penn State. Email: mxj222@psu.edu

†Assistant Professor, Dept. of Political Science & International Relations, University of Delaware. Email: bagozzib@udel.edu

‡Ph.D. Candidate, Dept. of Statistics, Penn State. Email: bzk147@psu.edu

§Professor, Dept. of Political Science, Penn State. Email: sxm73@psu.edu

Introduction

A remarkable amount of innovation has occurred within the study of survival (i.e., duration) models and related processes over the past several decades. One arena of political methodology insight in this regard relates to survival model extensions that seek to disentangle mixtures of multiple survival data processes. Box-Steffensmeier, De Boef and Joyce (2007), for example, introduce a survival model that separately accounts for within-observation heterogeneity arising from *both* event dependence *and* heterogeneity in repeated event processes; whereas Metzger and Jones (2016) introduce political scientists to a class of multi-state survival estimators that allow researchers to model distinct within-observation survival phases.[1] Others have extended the applicability of a class of split-population survival models known as cure models—which account for a type of "inflation" in one's non-failure survival cases that arises from the presence of observations that are effectively cured from ever experiencing an event failure of interest—for the study of Political Science (Box-Steffensmeier and Zorn, 1999; Svolik, 2008; Beger, Dorff and Ward, 2014, 2016); and with respect to the availability of corresponding open source software (Beger et al., 2017).

In this paper we contend that a reverse split-population survival process can also commonly arise within social science survival data. That is, survival data can often *over-report* events as having failed, such that *some* observations' true *censored* values are misclassified as *failed*. This leads to an inflation of failure events, whereas the cure model mentioned above instead accounts for inflation only within non-failure cases. Inaccurately recorded event failures of the former variety 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 terminated at that point in time is problematic as there is a non-zero probability that the observation persisted past that point. There are several social science scenarios where (a subset of) recorded failure events may actually persist be-

[1]Methodological research into non-proportional hazards is also illustrative in these regards (e.g., Keele, 2010; Licht, 2011; Jin and Boehmke, 2017; Ruhe, 2018).

yond their recorded failure time in this manner, leading to misclassification (i.e., inflation) in event failures. We discuss several such cases immediately below.

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 criteria 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.[2] We illustrate the consequences of these coding errors in the "Conservative Failure Threshold" subfigure below. Here, some cases persist beyond the window of time under analysis, and hence are accurately recorded as censored, whereas one remaining case is accurately recorded as failed. However, an additional subset of recorded failures in this subfigure persists beyond their recorded failure time, due to researchers' overly conservative thresholds for determining failures. Treating the latter cases as failures within 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.

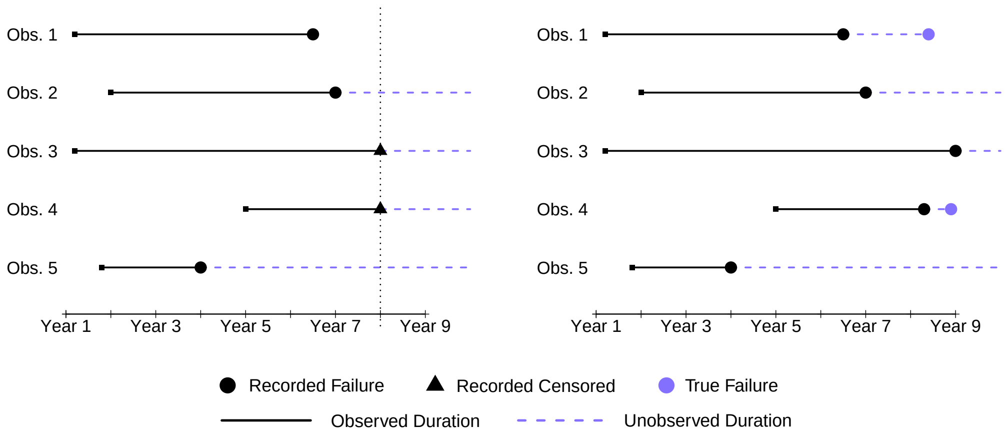
Misclassified failure events can also arise in survival data due to a variety of other coding or reporting processes. For example, within long-range historical analyses, studies of the

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

Figure 1: Misclassified Failure Illustrations

(a) Conservative Failure Threshold

(b) Last Observance Treated as Failure



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. In these cases, one's resultant survival data corresponds to the "Last Observance Treated as Failure" subfigure above: each observation's recorded failure time is an *underestimate* of that observation's true life-span, in that a researcher knows with certainty that that 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 in their severity with commonly studied covariates (e.g., environmental or geographic conditions), bias will again arise in survival estimates of these phenomenon. Finally, political actors often self-report their duration of (non)engagement in a given activity (e.g., political participation, compliance with a given law, or political donations), and these reports are often leveraged within survival analyses (e.g., Cress, McPherson and Rotolo, 1997; Box-Steffensmeier, Radcliffe and Bartels, 2005; Linos, 2007). For a subset of these cases, some actors may have strategic reasons to under-report their duration of (non)engagement.

Here again then, the recorded failures in one's survival data can exhibit misclassification, potentially in the manners depicted in either subfigure in Figure 1.

To address the methodological challenges associated with misclassified failures, we develop a a parametric misclassified failure split population 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. In a similar fashion to the cure survival model, our proposed model does so by estimating a system of two equations. The first can be characterized as a "splitting" equation that allows one to estimate the probability of a case being recorded as 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 (and corresponding coefficient estimates) are now estimated conditional on a case being a (non)misclassified failure instance. After deriving this model within both a non-time varying and time varying covariate context, we develop a corresponding R package to facilitate its estimation using a 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 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 misclassified failures are present, but also offers researchers with a means of theoretically identifying (and testing for) the factors that may govern a particular misclassified failure process.

1 Survival Model with Misclassified Failure

1.1 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 split population survival model for continuous time duration data, where subjects $i = \{1, 2, ...N\}$ 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 ($\widetilde{C}_i = 0$ 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 ($\widetilde{C}_i = 1$) provide information on both the hazard of an event, and the survival of individuals prior to that event, whereas censored observations ($\widetilde{C}_i = 0$) only provide information on an observation having survived at least until time T_i. Combining each set of observations' respective contributions to the density and survival functions, the likelihood and the log-likelihood function(s) of the standard parametric survival model are respectively,

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

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

where $\mathbf{X}_i$ are p1-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 $\widetilde{C}_i$ accurately records all censored cases $(\forall(\widetilde{C}_i = 0) : (C_i = 0))$ but mis-records some subset of non-censored failure outcomes as censored $(\exists\, (\widetilde{C}_i = 1) : (C_i = 0))$. 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 $\alpha_i = \Pr(\widetilde{C}_i = 1 | C_i = 0)$. 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:

$$\Pr(\alpha_i = 1) + \Pr(\alpha_i = 0) \Pr(t_i \leq T_i) = \alpha_i + (1 - \alpha_i) f(t_i), \tag{3}$$

with the corresponding unconditional survival function of

$$\Pr(\alpha_i = 0) \Pr(t_i > T_i) = (1 - \alpha_i) S(t_i), \tag{4}$$

where α_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:

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

where $\mathbf{Z}_i$ are p2-dimensional covariates and $\boldsymbol{\gamma}$ 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 α_i in (5), the log-likelihood function of the general paramteric split population model with misclassified failure cases (without time-varying covariates) is

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

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 $t0$ and "exit time" duration t for each period at which an observation is observed. As such, $t0_{ij}$ 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

$$S(t0) = 1 - F(t0), \tag{7}$$

where $F(t0) = \int_0^{t0} f(t0)$. With $S(t0)$ 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 $t0$:

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

As described in the Supplemental 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:

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

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 α_{ij} since the observed event failures may include latent misclassified failure 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 Supplemental Appendix – holds 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 of interest), the MF model with time-varying covariates is a model for instances where some subjects are observed as having failed or experienced the event of interest, 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 is contaminated with latent misclassified failure cases.

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 misclassified failure cases (e.g., civil conflict duration data) use the standard Weibull model, we develop and define the log-likelihood function of our MF *Weibull* model with time-varying covariates in the next section. In the Supplemental Appendix, we also develop the MF *exponential* model and assess this model via Monte Carlo simulations.

1.2 Misclassified Failure *Weibull* Model

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

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

In the Supplemental 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 MF *Weibull* model with time-varying covariates, which is given by:

$$
\begin{aligned}
&\ln L(\rho, \boldsymbol{\beta}, \boldsymbol{\gamma}) \\
&= \sum_{i=1}^{N} \left\{ \widetilde{C}_{ij} \ln \left[\alpha_{ij} + (1-\alpha_{ij}) \frac{\exp(\mathbf{X}_{ij}\boldsymbol{\beta})\rho\ (\exp(\mathbf{X}_{ij}\boldsymbol{\beta})t_{ij})^{\rho-1} \exp((-\exp(\mathbf{X}_{ij}\boldsymbol{\beta})t_{ij})^{\rho})}{\exp(-(\exp(\mathbf{X}_{ij}\boldsymbol{\beta})t0_{ij})^{\rho})} \right] \right. \\
&\qquad \left. + (1-\widetilde{C}_{ij}) \ln \left[(1-\alpha_{ij}) \frac{\exp(-(\exp(\mathbf{X}_{ij}\boldsymbol{\beta})t_{ij})^{\rho})}{\exp(-(\exp(\mathbf{X}_{ij}\boldsymbol{\beta})t0_{ij})^{\rho})} \right] \right\}.
\end{aligned} \tag{11}
$$

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,[3] we estimate this model via the 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 – ρ, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ – and then define the conditional posterior distribution of these parameters. Following standard practice, we assign the multivariate Normal prior to $\boldsymbol{\beta} = \{\beta_1, ..., \beta_{p_1}\}$ and $\boldsymbol{\gamma} = \{\gamma_1, ..., \gamma_{p_2}\}$, and the Gamma prior for ρ with shape and scale parameters a_ρ and b_ρ:

$$
\begin{aligned}
\rho \quad \sim \quad \text{Gamma}(a_\rho, b_\rho), \quad \boldsymbol{\beta} \sim \text{MVN}_{p_1}(\mathbf{0}, \varSigma_\beta), \quad \boldsymbol{\gamma} \sim \text{MVN}_{p_2}(\mathbf{0}, \varSigma_\gamma) \\
\varSigma_\beta \sim \text{IW}(S_\beta, \nu_\beta) \qquad \varSigma_\gamma \sim \text{IW}(S_\gamma, \nu_\gamma),
\end{aligned} \tag{12}
$$

where a_ρ, b_ρ, S_β, ν_β, S_γ, ν_γ are the hyperparameters. Note that we use hierarchical Bayesian modeling to estimate $\varSigma_\beta$ and $\varSigma_\gamma$ using the Inverse-Wishart (IW) distribution. Given these prior specifications and the hyperparameters, the conditional posterior distributions for ρ, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ parameters in the Bayesian MF Weibull model (with time-varying covariates) are

[3] The Broyden, Fletcher, Goldfarb, Shannon (BFGS) method in the R *optim* function. In our Monte Carlo analysis, we briefly assess the properties of the MF Weibull model estimated by BFGS.

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

where $P(\mathbf{C},\mathbf{X},\mathbf{Z},\mathbf{t},\mathbf{t0},\boldsymbol{\beta},\boldsymbol{\gamma},\rho)$ is the likelihood that can be obtained using the log-likelihood in equation (11), and $P(\rho|a_\rho,b_\rho)$, $P(\boldsymbol{\beta}|\Sigma_\beta)$, and $P(\boldsymbol{\gamma}|\Sigma_\gamma)$ 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 ρ, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ are not available, we use MCMC methods with the following slice sampling (Neal, 2003) update scheme,

- **Step 0.** Choose initial value of $\boldsymbol{\beta}, \boldsymbol{\gamma}$, and ρ and set $i = 0$.
- **Step 1.** Update $\Sigma_\beta \sim P(\Sigma_\beta|\boldsymbol{\beta})$ and $\Sigma_\gamma \sim P(\Sigma_\gamma|\boldsymbol{\gamma})$ from conjugate posteriors. The closed form of the full conditional distributions for Σ_β and Σ_γ are derived in the Supplemental Appendix.
- **Step 2.** Update $\boldsymbol{\beta} \sim P(\boldsymbol{\beta}|\mathbf{C},\mathbf{X},\mathbf{Z},\mathbf{t},\mathbf{t0},\boldsymbol{\gamma},\rho,\Sigma_\beta)$, $\gamma \sim P(\boldsymbol{\gamma}|\mathbf{C},\mathbf{X},\mathbf{Z},\mathbf{t},\mathbf{t0},\boldsymbol{\beta},\rho,\Sigma_\gamma)$ and $\rho \sim P(\rho|\mathbf{C},\mathbf{X},\mathbf{Z},\mathbf{t},\mathbf{t0},\boldsymbol{\beta},\boldsymbol{\gamma},a_\rho,b_\rho)$ using slice sampling. We use the univariate slice sampler with stepout and shrinkage (Neal, 2003). Detailed steps to perform slice sampling for $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and ρ are described in the Supplemental 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 (via, e.g., credible intervals or posterior means).

2 Monte Carlos

We conduct 11 Monte Carlo (MC) experiments to assess the relative performance of the survival models discussed above. Our primary MC experiments simulate either (i) a non-MF Weibull distributed outcome variable (Experiment 1) or (ii) 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, under circumstances where $N = 1,000$, $N = 1,500$, or $N = 2,000$. Experiments 3-4 instead assess the performance of 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; again for $N = 1,000$, $N = 1,500$, and $N = 2,000$. Experiments 5-8 simulate an exponentially distributed[4] outcome variable (Experiments 5 and 7), or a MF exponential outcome variable (Experiments 6 and 8), and compare the relative 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). As above, Experiments 5-8 evaluate all models considered under conditions of $N = 1,000$, $N = 1,500$, and $N = 2,000$. Finally, Experiments 9-11 return to our Bayesian Weibull and Bayesian MF Weibull models and compare these two estimators under instances of increasingly larger MF rates; again for our three N's of interest.

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 MF outcome experiments (i.e., Experiments 2, 4, 6, 8, and 9-11) then add a moderate level of misclassified failure 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, respectively). To generate our MF rates in Experiments 2, 4, 6, 8, and 9-11, we define a set of misclassification stage covariates $\mathbf{z} = (\mathbf{1}, \mathbf{z_1}, \mathbf{z_2})'$,where $\mathbf{z_1} = ln(Uniform[0, 100])$ 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 (Experiments 1-11). Our misclassification stage parameters are defined as $(\gamma_1, \gamma_2, \gamma_3)' = (-2, 3, 3)'$ (Experiments 2, 4, 6, and 8), 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) use $\rho = 2$. For each parameter estimate, we retain and evaluate the mean (MCMC-simulated) estimate and (MCMC-simulated) standard error (MCSE), as well as the parameter estimate's root mean square error (RMSE).

[4] A Weibull distributed outcome variable with $\rho = 1$.

Experiment 1 evaluates the relative performance of (i) a Bayesian Weibull model and (ii) a Bayesian MF Weibull model when the true data generating process (d.g.p.) corresponds to a Weibull survival process with no misclassified failures. 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 Supplemental 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 β parameters of interest, the Bayesian Weibull and Bayesian MF Weibull models recover averaged parameter estimates that are virtually identical. This is corroborated by the RMSEs 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. Indeed, there are several instances where the Bayesian MF Weibull model exhibits slightly less bias than the Bayesian Weibull (e.g., for β_0 when $N = 1,000$ or $N = 1,500$). However, although both models consistently exhibit low MCSEs, the Bayesian Weibull model's MCSEs are consistently smaller than those of the Bayesian MF Weibull.

Table 1: Markov Chain Monte Carlo (MCMC) β-Estimates for Experiments 1 and 2

	Experiment 1: Non-MF Weibull D.G.P.						
#Obs.	Model	$\hat{\beta}_0$	$MCSE(\hat{\beta}_0)$	$RMSE(\hat{\beta}_0)$	$\hat{\beta}_1$	$MCSE(\hat{\beta}_1)$	$RMSE(\hat{\beta}_1)$
1,000	Bayes Weibull	0.999	9.95E-05	0.027	3.500	1.37E-05	0.004
	Bayes MF Weibull	1.002	7.01E-04	0.025	3.500	9.07E-05	0.004
1,500	Bayes Weibull	1.002	6.69E-05	0.023	3.500	9.31E-06	0.003
	Bayes MF Weibull	1.001	4.47E-04	0.022	3.500	5.80E-05	0.003
2,000	Bayes Weibull	1.002	5.11E-05	0.019	3.500	7.10E-06	0.003
	Bayes MF Weibull	0.999	3.25E-04	0.020	3.500	4.23E-05	0.003
	Experiment 2: MF Weibull D.G.P.						
#Obs.	Model	$\hat{\beta}_0$	$MCSE(\hat{\beta}_0)$	$RMSE(\hat{\beta}_0)$	$\hat{\beta}_1$	$MCSE(\hat{\beta}_1)$	$RMSE(\hat{\beta}_1)$
1,000	Bayes Weibull	1.226	1.66E-04	0.226	3.478	2.42E-05	0.022
	Bayes MF Weibull	1.005	2.30E-04	0.026	3.499	2.57E-05	0.004
1,500	Bayes Weibull	1.237	1.14E-04	0.237	3.476	1.68E-05	0.024
	Bayes MF Weibull	1.002	1.20E-04	0.022	3.500	1.41E-05	0.003
2,000	Bayes Weibull	1.248	8.87E-05	0.248	3.475	1.29E-05	0.025
	Bayes MF Weibull	1.002	3.17E-04	0.019	3.500	3.15E-05	0.003

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

In sum, while the Bayesian Weibull model outperforms the Bayesian MF Weibull model

Table 2: Markov Chain Monte Carlo (MCMC) γ-Estimates for Experiment 2

	Experiment 2: MF Weibull D.G.P.									
#Obs.	Model	$\hat{\gamma}_0$	$MCSE(\hat{\gamma}_0)$	$RMSE(\hat{\gamma}_0)$	$\hat{\gamma}_1$	$MCSE(\hat{\gamma}_1)$	$RMSE(\hat{\gamma}_1)$	$\hat{\gamma}_2$	$MCSE(\hat{\gamma}_2)$	$RMSE(\hat{\gamma}_2)$
1,000	Bayes MF Weibull	-1.584	0.147	0.838	1.959	0.108	0.344	3.285	0.141	0.532
1,500	Bayes MF Weibull	-1.697	0.113	0.797	1.935	0.069	0.305	3.290	0.083	0.443
2,000	Bayes MF Weibull	-1.628	0.341	0.705	2.032	0.249	0.208	3.272	0.327	0.378

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

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 (β parameters) and in Table 2 (γ parameters). We also plot the full distributions of each β parameter in the Supplemental 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 β 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. By contrast, the standard Bayesian Weibull model's mean $\hat{\beta}$'s substantially overestimate β_0 and typically underestimate β_1 no matter the N considered. These conclusions are reinforced by the RMSE and MCSE values reported in Table 1, which indicate that the Bayesian MF Weibull model exhibits RMSE values that are 2-10 times smaller than the standard Bayesian Weibull models' RMSEs, and exhibits MCSEs that are generally comparable in size to those of the Bayesian Weibull.

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 next to Experiment 2's MF Weibull γ estimates (Table 2), we find in our averaged $\hat{\gamma}$ values that our Bayesian MF Weibull model generally recovers each true γ value quite well. That being said, the RMSE and MCSE values reported in Table 2 nevertheless suggest that the Bayesian MF Weibull model's $\hat{\gamma}$'s exhibit higher bias, and lower efficiency, than was the case for the Bayesian MF Weibull's $\hat{\beta}$'s in Experiment 2. This disparity declines as one increases N from 1,000 to 2,000.

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 Supplemental 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 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 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, we can again conclude that the risks to inference of (mis)applying the MF Weibull in the absence of misclassified failures are fairly low, whereas the inferential risks of (mis)applying a standard 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 the each MF model are comparable across Experiments 2 and 4, as are the corresponding RMSEs. However, when one's outcome variable is MF Weibull, we also observe that the Bayesian MF Weibull model's $\hat{\gamma}$'s generally exhibit lower bias, and higher efficiency, than do those of the BFGS MF Weibull. Thus, we can conclude from Experiments 1-4 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 Supplemental 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 misclassified failures. 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 and efficiency and (ii) the Bayesian MF survival models 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.

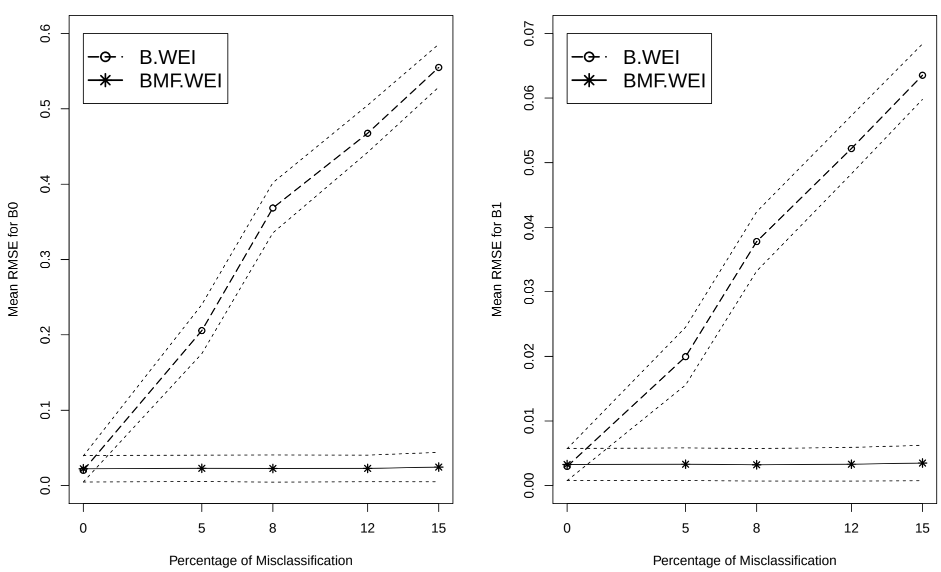
Whereas Experiments 2, 4, 6, and 8 employ a MF rate of 5%,[5] Experiments 9-11 noticeably increase this MF rate above 5%. These latter experiments—which we present in the Supplemental Appendix—increasingly favor the Bayesian MF Weibull model over the Bayesian Weibull model as one's MF rate extends beyond 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 ($\alpha = 8\%$), Experiment 10 ($\alpha = 10\%$), and Experiment 12 ($\alpha = 15\%$). 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 ($\hat{\beta}_1$). These Figures demonstrate that 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

[5] Which we suspect will be most comparable to the types of applications considered below.

increases, we find that the Bayesian MF Weibull model's $\hat{\beta}$ RMSEs remain effectively flat, whereas those the Bayesian Weibull dramatically increase. Herein, the Bayesian MF Weibull already exhibits RMSE's that are over 7 times smaller than those of Bayesian Weibull when $\alpha = 5\%$, and these Bayesian MF Weibull RMSEs then become nearly 20 times smaller than those of the Bayesian Weibull when α is increased to 15%. Hence, Experiments 9-11 further underscore the preferability of the Bayesian MF Weibull in situations of modest-to-moderate misclassified failures.

Figure 2: Comparison of Survival Stage RMSEs Under Increasing Misclassified Failure Rates

(a) Averaged RMSEs for $\hat{\beta}_0$ (b) Averaged RMSEs for $\hat{\beta}_1$



3 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. For our first application, 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 log distance from the civil conflict center to the capital city (*distance to capital* (*ln*)) and civil conflicts in border regions (*conflict at border* dummy) decreases the hazard of civil war termination or equivalently leads to longer civil wars, while higher *rebel fighting capacity* increases the hazard of civil war failure. Alongside these assessments, they empirically analyze the effect of the following three covariates on civil war duration that are studied in the theoretical literature on civil wars. First, following Collier, Hoeffler and Soderbom (2004) and Fearon and Laitin's (2003) theoretical claim, Buhaug *et al* assess whether countries with higher GDP per capita at the onset of civil wars are likely to be associated with a higher hazard of civil war termination (i.e., shorter civil conflicts). They also test – as suggested by Balcells and Kalyvas (2014) and Straus (2012) – whether civil war duration has declined following the end of the Cold War which implies that the post-Cold War era is associated with a higher hazard of civil war failure. Furthermore, following Cunningham (2006) and Thyne's (2012) research on domestic institutions and civil conflict duration, Buhaug *et al* (2009: 551-554, 563) test whether "higher democracy scores" at the onset of civil wars are associated with longer civil conflicts.

To statistically assess these theoretical predictions, Buhaug *et al* 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), Buhaug *et al* (2009) operationalize the termination of a civil conflict according to its official end date (or "failure"), which is based on 24-month spells where the UCDP/PRIO ACD recorded fewer than 25 battle-deaths per year. That is, civil conflict is coded as "terminated" in the Buhaug *et al* data when the number of battle-deaths *falls and stays* below 25 for at least 24 months. Their civil war duration data have a mean of 2,221 days with a total of 149 civil conflict termination episodes (Buhaug *et al* 2009: 556).

Buhaug *et al* estimate a standard parametric MLE Weibull model on their survival data in which they include the following covariates listed above that separately assess how geography and the fighting capacity of rebel groups influence civil war duration: *distance to capital*(*ln*), *conflict at border*, and a binary measure of *rebel* (group) *fighting capacity*. Drawing on extant research, they also incorporate the following additional covariates: *GDP capita at onset* (*ln*) of civil wars, a dummy for *post Cold War* years, a measure of *democracy score at onset* of civil wars, 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* and *distance to capital*(*ln*) have a reliably negative (positive) influence on the hazard of civil war failure (civil war duration). Additionally, they report that the statistical association between *GDP capita at onset* (*ln*) and the hazard of civil conflict termination are positive, but unreliable. They, however, find robust support for the theoretical claims that *post Cold War* and *rebel fighting capacity* have a statistically positive and reliable effect on the hazard of civil war failure therein implying that these covariates are each associated with shorter civil wars. *Democracy score at onset* has a statistically negative and highly reliable effect on the hazard of civil conflict termination in their MLE Weibull model, as anticipated by Buhaug *et al* (2009: 563)

Although now standard practice in the civil war literature (Themnér and Wallensteen, 2014; Thyne, 2012), 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 misclassified failure cases in data. First, the use of 24-month spells to identify conflict termination is arguably *conservative*, especially for lower intensity conflicts in remote or poor information environments, or in situations where some groups or officials do not recognize the war as having ended. Such cases where the date of civil conflict termination is ambiguous are unlikely to capture the "true" termination date, and several entities might argue that conflict ended in different periods. Take, for instance, the Second Congo War, which officially ended in 2003. Is the correct termination date July 2003, the month and year 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 despite the fact that other key sources (Coghlan et al., 2009) identify battle deaths to be in the thousands? Or maybe the war never ended, considering that conflict in the

Democratic Republic of Congo (DRC) still takes the lives of thousands every year (Larmer, Laudati and Clark, 2013). Without perfect information, the dates used to record civil war termination are likely to be underestimated at times, leading to misclassified failure cases in civil war termination datasets such as Buhaug *et al's* data.

A second issue is the possibility that different sources may record distinct dates for civil war termination even though they use the same criterion (e.g., battle-death numbers threshold) to code the "end" of civil wars. 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 key sources that use the same UCDP battle-death threshold criterion emphasize that civil conflict between the Indian Government and Nagaland's rebel groups during the 1990s did not "end" in 1997 but rather persisted into the first decade of the twenty-first century (Shimray 2001). This discrepancy is not surprising given the subjectivity involved in accurately identifying the number of battle deaths required to code civil war termination. The Nagaland example in the Buhaug *et al* data is not rare. Indeed, Table A.9 in the Supplemental Appendix identifies many additional terminated civil conflict cases in the Buhaug *et al* data—including civil wars in other parts of India, Myanmar, the DRC, and Thailand—that persisted beyond their recorded time. These examples thus suggest that civil war duration datasets including Buhaug *et al's data* are likely contaminated with misclassified failure cases that have persisted beyond their observed failure point.

We therefore 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 misclassified failure cases in Buhaug *et al's* civil war duration data, the (ii) standard MLE Weibull model and (iii) a standard Bayesian Weibull model. Table 3 reports the results from these models that focuses on the Buhaug *et al* specification mentioned earlier which evaluates the effect of the following variables on *civil war duration*: *GDP capita at onset* (*ln*), the *post Cold War* dummy, *Democracy score at onset*, *rebel fighting capacity*, *distance to capital*(*ln*), *conflict at border*, and *Border* × *distance (ln)*. For the first two models in Table 3, we estimate a standard Weibull hazard model first via MLE (whose coefficient estimates are reported in Model 1) and then via Bayesian MCMC (whose posterior mean estimates are presented in Model 2) to assess the aforementioned specification. Models 3-6 in Table 3 report the posterior mean estimates with 95% Bayesian Credible Intervals (hereafter, BCI) obtained from different specifications for four different Bayesian MF Weibull models. Importantly, recall here that, unlike the standard Weibull model, the Bayesian MF Weibull estimates the effect of both (i) a series of **X** covariates on *civil war duration*, and (ii) a set of **Z** covariates on the probability of failure misclassification (denoted α).

We thus first report a baseline Bayesian MF Weibull specification in Model 3 of Table 3. The survival stage in this baseline MF model of *civil war 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 Bayesian MF Weibull's survival stage in Model 4 of Table 3 repeats the survival stage specification outlined above, but adds a set of 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 attention with respect to civil war coverage (Collier 2003, Puddephatt 2006). This improves the accuracy of information about civil war 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) tends to be inaccurate in civil conflicts fought in remote geographic areas that are far away from the capital city (Puddephatt 2006). This covariate is thus likely to be positive in the misclassification stage. We also incorporate *conflict at border* as researchers argue that governments in civil war-affected countries tend to misrepresent information about battle-related deaths in civil conflicts in their state's border regions to demonstrate that government forces are "winning" the civil war (Buhaug and Gates 2002, Lischer 2015). *Conflict at border* is hence likely to be positive in the misclassification stage as information about battle-related deaths might be inaccurate in civil wars that occur in the border zones of conflict-affected states.

The Bayesian MF Weibull's survival stage in Model 5 (Table 3) also repeats the survival stage specification used in Buhaug *et al* (2009), while the model's misclassification stage includes the three covariates discussed above and *rebel fighting capacity*. Finally, Model 6 in Table 3 includes all covariates from the Buhaug *et al* specification for *both* the survival stage and the misclassification stage of the Bayesian MF Weibull specification. We use the slice-sampling (MCMC) algorithm described earlier to estimate all the Bayesian MF Weibull models, and to this end, we specify the hyperparameters as: $a = 1$, $b = 1$, $S_\beta = I_{p1}$, $S_\gamma = I_{p2}$, $\nu_\beta = p1$ and $\nu_\gamma = p2$.[6] We first discuss the Bayesian MF Webiull model's misclassification stage and then the model's survival stage results. The densities (Figure A.12, Supplemental Appendix) and posterior mean estimates of the misclassification stage covariates show that *conflict at border* is consistently positive in the Bayesian MF Weibull's misclassification stage in Models 4-6 and this estimate's 95% Bayesian Credible Intervals (hereafter, BCI) always excludes zero. The first difference in misclassification probabilities derived from the Bayesian MF Weibull model's misclassification stage (**Z**) covariates in Model 4—illustrated in Figure 3 with 95% BCI—further reveal 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 5.86%. The 95% BCI of this effect excludes zero which means that it is (as predicted theoretically) 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. In line with our theoretical expectations, the posterior mean estimate and substantive effect of *distance to capital* (*ln*) in the misclassification stage (see Figure 3) shows that civil conflicts fought in

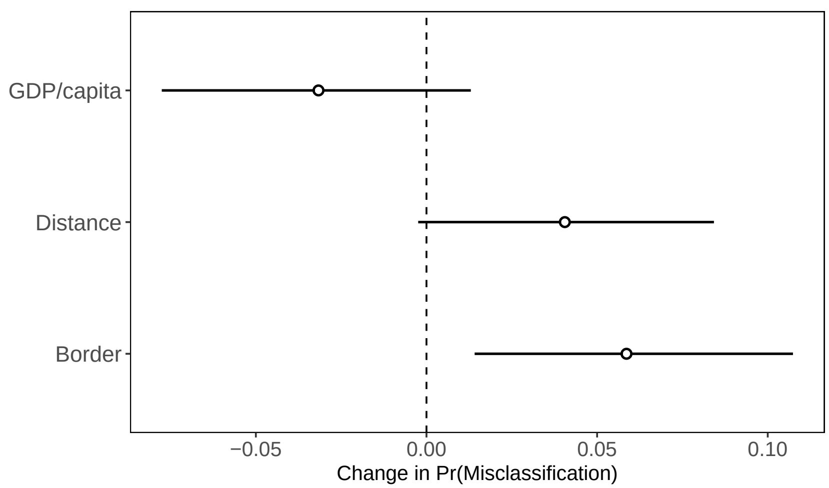
[6]The Bayesian MF Weibull model's results in Table 3 are based on a set of 50,000 iterations after 4,000 burn-in scans and thinning of 10.

Table 3: (Bayesian MF) Weibull Results: Buhaug *et al* (2009) Application

	MLE Weibull	Bayesian non-MF Weibull	Bayesian MF Weibull			
	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)
X *Covariates* ($\boldsymbol{\beta}$ *Parameters*)						
Distance to capital (ln)	-0.325 (0.084)	-0.582 (-0.81, -0.327)	-4.586 (-7.686, -1.472)	-2.427 (-7.537, 3.082)	-4.971 (-10.136, -0.762)	-3.744 (-8.177, 1.825)
Conflict at border	-0.468 (0.200)	-0.727 (-1.27, -0.159)	-0.483 (-5.751, 3.548)	-1.923 (-10.627, 4.384)	-3.831 (0.063, -9.865)	-3.272 (-11.524, 2.120)
Border×Distance	0.459 (0.115)	0.769 (0.4, 1.123)	1.771 (-4.787, 8.496)	1.256 (-4.756, 9.389)	0.86 (6.120, -2.023)	4.796 (-3.414, 11.308)
Rebel fighting capacity	0.359 (0.183)	0.425 (-0.118, 0.953)	4.11 (-2.684, 8.507)	-0.201 (-9.018, 6.764)	1.712 (4.499, -1.570)	6.136 (-0.402, 9.441)
Democracy score at onset	-0.588 (0.222)	-0.921 (-1.62, -0.247)	-1.711 (-9.567, 5.053)	-0.614 (-12.377, 7.225)	1.445 (7.201, -2.230)	-6.789 (-10.635, -1.687)
GDP capita at onset (ln)	0.115 (0.081)	0.047 (-0.194, 0.318)	-7.85 (-12.313, -1.103)	-4.745 (-17.375, -0.544)	-3.161 (-8.031, -0.102)	-6.078 (-11.477, -1.590)
Post-Cold War years	0.547 (0.184)	0.903 (0.382, 1.438)	-4.05 (-11.011, 1.487)	-0.03 (-8.818, 7.181)	-1.228 (-3.87, 1.212)	-4.66 (1.261, -13.757)
Constant	-3.839 (0.777)	-4.407 (-6.83, -2.315)	-1.856 (-5.436, 1.009)	-0.103 (-7.778, 7.449)	0.522 (-2.011, 3.33)	5.986 (-1.402, 13.605)
Z *Covariates* ($\boldsymbol{\gamma}$ *Parameters*)						
Distance to capital (ln)				0.124 (-0.007, 0.248)	0.068 (-0.071, 0.191)	0.385 (0.190, 0.579)
Conflict at border				0.495 (0.131, 0.86)	0.484 (0.102, 0.858)	0.568 (0.131, 0.969)
Border×Distance						-0.582 (-0.874, -0.303)
Rebel fighting capacity					-0.538 (-0.903, -0.184)	-0.388 (-0.751, -0.03)
Democracy score at onset						0.681 (0.187, 1.185)
GDP capita at onst (ln)				-0.093 (-0.214, 0.039)	-0.748 (-1.081, -0.395)	-0.101 (-0.253, 0.052)
Post-Cold War years						-0.7 (-1.096, -0.329)
Constant			1.974 (1.807, 2.146)	1.533 (0.382, 2.751)	1.728 (0.885, 2.648)	0.242 (-1.132, 1.657)
DIC		-23764.5	-13905.15	-18339.6	-19301.91	-7211.451
AIC	724.779					
log likelihood	-353.4	581.98	680.531	703.746	629.007	602.347
Observations	1,375	1,375	1,375	1,375	1,375	1,375

Note: Variable coefficients are reported with standard errors clustered by conflict in parentheses for the MLE Weibull model. Posterior means are reported with 95% credible intervals in parentheses for the Bayesian non-MF and MF Weibull models.

Figure 3: Change in the Predicted Probability of Misclassification



geographically remote areas are indeed more likely to be misclassified as having failed when they had not. However, the 95% BCI of this covariate's mean estimate and substantive effect in the misclassification stage always includes zero thus indicating that the aforementioned empirical relationship is unreliable.

In addition, the posterior densities and mean estimate of the misclassification stage covariates show that *GDP capita at onset* (*ln*) is consistently negative in the Bayesian MF Weibull's misclassification stage in Models 4-6. The 95% BCI of this covariate's estimate excludes zero in some—but includes zero in other—misclassification stage specifications. Further, Figure 3 shows that increasing *GDP capita at onset* (*ln*) from 1 standard deviation (SD) below to 1 SD above its mean decreases the probability of a misclassified war failure by approximately 3.16%, although the 95% BCI of this effect includes zero. While this supports our claim that civil wars 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 not reliable. Intuitively, other misclassification stage results show that *rebel fighting capacity* and *post Cold War* dummy are also negatively associated with the probability of misclassification failure.

We next turn to the (MF) Weibull survival stage results from Table 3. First, the results of the covariates in the standard MLE and Bayesian non-MF Weibull models are not only similar but also confirm all the results that Buhaug *et al* (2009) report. For instance, the influence of the following three covariates – *Distance to capital*(*ln*), *Conflict at border* and *Democracy score at onset* – on the hazard of civil conflict termination is negative and highly reliable in the standard MLE and Bayesian Weibull models, which mirrors Buhaug *et al's* findings. The estimate of *rebel fighting capacity* increases the hazard of civil war failure reliably 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 model but statistically unreliable. This is identical to Buhaug *et al* (2009: 563), who find that although unreliable, higher per capita income at the onset of civil wars is indeed associated with a higher hazard of civil conflict termination. The positive estimate of the *Post-Cold*

War dummy is statistically reliable in the MLE and Bayesian non-MF Weibull models. This indicates that the hazard of civil war failure has increased or, in other words, the duration of civil wars has declined in the post Cold War era which is exactly what Buhaug *et al* (2009) and Balcells and Kalyvas (2014) find.

However, the Bayesian MF Weibull's estimates differ substantially from the standard Bayesian non-MF and MLE Weibull model's that Buhaug *et al* (2009) report. To see this, we focus on the (i) top rows of Models 3-6 in Table 3, (ii) hazard ratio plots of the Bayesian MF Weibull model's key survival stage covariates in Figures 4a-4b and (iii) the survival stage covariates' posterior densities in Figure A.13 in the Supplemental Appendix. Herein, the density and posterior mean survival stage estimate of both *distance to capital* (*ln*) and *conflict at border* reveals that each of these two covariates are negatively associated with the hazard of civil war failure in the MF Weibull models, although this association is unreliable since the 95% BCIs of these variables frequently 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 civil war failure is highly robust in their MLE Weibull model.

Next, the density and posterior mean survival stage estimate of log of *GDP capita at onset* in the Bayesian MF Weibull specification in Model 3 (where the misclassification stage only includes the intercept) is negative, specifically -7.85, with a 95% BCI range of [-12.313, -1.103] that excludes zero. The survival stage estimate of *GDP capita (onset)* remains negative and its 95% BCI *always* excludes zero in the remaining Bayesian MF Weibull models 4-6 in Table 3 in which the misclassification stage includes covariates. Hence per capita income at the onset of civil wars has a reliably negative influence on the hazard of civil war failure in the Bayesian MF Weibull survival stage, which is exactly the opposite of what Buhaug *et al* and we find in the standard MLE and Bayesian Weibull model. This suggests that a possible *prolonging* effect of GDP per capita on civil war duration may have gone unnoticed in many past analyses,[7] which failed to take into account misclassified failure cases (and

[7]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 which is distinct from

these misclassified failure cases' potential association with low GDP per capita).

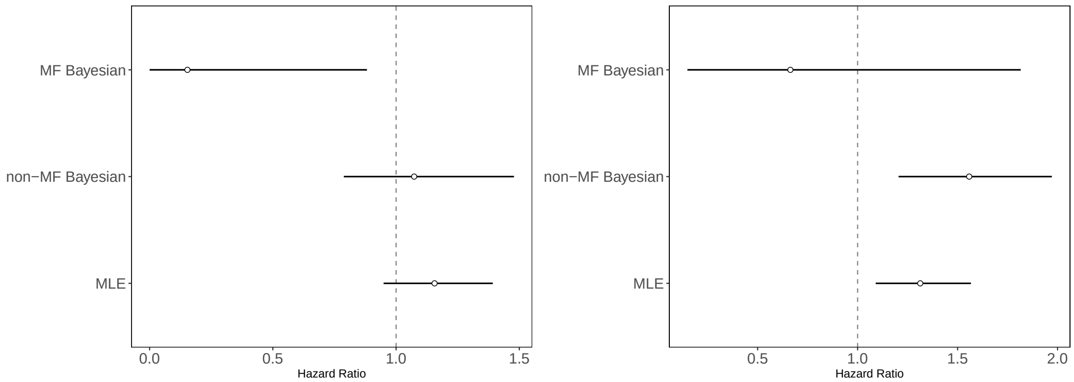
Additionally, consider Figure 4a. This figure illustrates hazard ratio plots derived from the estimate of *GDP capita at onset* (*ln*) in the (i) Bayesian MF Weibull specification in Model 4, (ii) standard MLE Weibull specification in Model 1, and (iii) standard Bayesian Weibull specification in Model 2. Note that a hazard ratio greater (lesser) than one suggests that this variable increases (decreases) the hazard of civil war termination.We learn from Figure 4a 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 respective mean, <u>increases</u> the hazard of civil conflict termination in the standard MLE and Bayesian non-MF Weibull models. But this effect is unreliable the MLE and Bayesian non-MF Weibull models. Thus, while there exists a positive association between economic development and the hazard of civil war failure – as suggested theoretically by Collier, Hoeffler and Söderbom (2004) – this association is tenuous. In sharp contrast, as shown in Figure 4a, increasing *GDP capita at onset* (*ln*) from 1 SD below to 1 SD above its mean <u>decreases</u> the hazard of civil conflict termination by 84.6% in the Bayesian MF Weibull model and the 95% BCI of this hazard ratio excludes zero. Hence, although reasonable increases in per capita income at the outbreak of civil wars increases the hazard of civil war failure in the standard Weibull models, the same changes in *GDP capita at onset* (*ln*) in the Bayesian MF Weibull specification leads to a substantial and reliable decrease in the hazard of civil war termination after misclassified failures are accounted for.

We turn to analyze another key variable that Buhaug *et al* (2009) evaluate, namely the survival stage estimate of *post Cold War* years. The hazard ratio plot in Figure 4b shows that increasing the *post Cold War* dummy from 0 to 1 while holding the other survival stage covariates at their mean increases the hazard of civil conflict termination substantially and reliably in the standard MLE Weibull and the non-MF Bayesian Weibull models. This finding corroborates Balcells and Kalyvas' (2014) theoretical claim and Buhaug *et al's* (2009)

our robust finding for this covariate), as well as Brandt *et al* (2008), who find that GDP per capita increases civil war duration for only civil wars ending in government victory specifically.

Figure 4: Hazard Ratios for GDP/capita at onset and Post-Cold War

(a) Hazard Ratio for GDP/capita at onset (b) Hazard Ratio for Post-Cold War



finding that the hazard of civil war failure has increased in the post cold-war period or equivalently, that the duration of civil wars has declined in the post cold-war era. But in sharp contrast to the standard Weibull model's results for this variable, the posterior mean estimate of *post Cold War* is negative in the survival stage of all the Bayesian MF Weibull specifications. Figure 4b shows that increasing the *post Cold War* dummy from 0 to 1 while holding the other survival stage covariates at their mean decreases the hazard of civil war failure by 33.63% in the Bayesian MF Weibull specification; however, the 95% BCI of this effect includes zero. Table 3 reveals similar contradictory results for several additional key survival stage covariates from the Buhaug *et al* (2009) study, including *democracy score at onset* and *rebel fighting capacity*—suggesting that past findings in these regards may have at least been partly attributable to misclassification in civil war failures, and to associations between failure misclassification and the variables reviewed here.

Importantly, Deviance Information Criterion (DIC) – which are discussed in more detail in the Supplemental Appendix – also suggest that these alternate Bayesian MF Weibull results are in fact preferable (in terms of overall fit) to those obtained from the Bayesian Weibull model reported in Table 3. Moreover, as reported in Table A.10 in the Supplemental Appendix, the posterior mean estimates of all the key survival stage covariates in the Bayesian MF Weibull models such as per capita income, post cold war years, and democracy score in models 3-6 in Table 3 remain robust in an extensive array of additional specifications that includes alternative controls in the MF Weibull model's misclassification stage. Standard diagnostic checks for the parameters in each Bayesian MF Weibull specification in Table 3 further suggest that the Markov chain has reached a steady state in each case. For instance, the trace plots show that the Markov chain has stabilized, has good mixing and is dense. Autocorrelation plots indicate no high degree of autocorrelation for the posterior samples, implying good mixing.[8] Hence the plots suggest that the Markov chain has converged to the desired posterior.

Altogether, comparisons of Table 3's 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 misclassified failures 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. These findings suggest that more attention should be paid to the "fuzziness" of civil war termination dates in empirical conflict research, and that the MF models proposed above may allow for one means for analysts to do so. Moreover, the empirical findings from the misclassification stages of the Bayesian MF Weibull models that we discuss above suggest that the MF hazard models also allow researchers to assess when failure cases in survival datasets are more likely to be misclassified, which is both substantively appealing and empirically useful.

Our second application is presented in full in the Supplemental Appendix and focuses on Reenock, Bernhard and Sobek's (2007) (hereafter RBS) study of democratic regime survival for the years 1961-1995. RBS posit that higher levels of deprivation of basic civilian needs (i.e. food insecurity) increase the prospects of democratic regime breakdown when per capita income within a given country reaches a certain threshold. They evaluate this moderation effect by interacting their explanatory variable, *basic needs deprivation*, with *GDP per capita (logged)* in a standard MLE Weibull model whose outcome is the duration of democratic

[8] Geweke-Diagnostics also show that the mean estimate of the Markov chain is stable over time.

regimes (*democratic survival*). RBS find that the statistical association between *basic needs deprivation* × *GDP per capita* (*logged*) and *democratic survival* strongly supports their expectations. They also find that increasing *basic needs deprivation* from 1 SD below to 1 SD above its mean in democracies reliably increases the hazard of democratic regime "failure" when per capita income reaches $2,300 in their sample (RBS 2007, 692).

We contend in the Supplemental Appendix that the criteria that RBS use to code democratic breakdowns likely means that their data contain some misclassified democratic-regime failure cases. Indeed, Table A.11 lists numerous examples of recorded democratic-regime failure in the RBS data that are ambiguous and hence likely misidentified. We thus replicate RBS' main specification with both a standard non-MF Weibull model and different specifications of our Bayesian MF Weibull model. Following RBS, the survival stage of the latter models includes *basic needs deprivation* × *GDP per capita* (*logged*), the individual components of this interaction term, and a set of controls that affect *democratic survival.* The misclassification stage covariates (and their results) in the Bayesian MF Weibull models applied to the RBS (2007) data are presented in Table A.12 and Figures A.14-A.15, while the survival stage results from these MF models are presented in Table A.12 and Figure A.16 in the Supplemental Appendix.

In brief, the standard Weibull's results are identical to RBS' finding as *basic needs deprivation* × *GDP per capita* (*logged*) increases the hazard of democratic regime failure in these models. In sharp contrast, the posterior mean survival stage estimate of *basic needs deprivation* × *GDP per capita* (*logged*) in the Bayesian MF Weibull models suggest that this interaction term decreases the hazard of democratic regime failure although the estimate's 95% BCI includes zero. The marginal effect of this interaction term derived from the main Bayesian MF Weibull survival stage shows that increasing *basic needs deprivation* from 1 SD below to 1 SD above its mean increases the duration (i.e. decreases the hazard) of democratic regimes by 24.5% when per capita income reaches the $2,300 threshold (the RBS benchmark) and the 95% BCI of this effect excludes zero. Thus, after statistically accounting for misclassified democratic regime-failure cases, the association between basic needs deprivation and democratic regime failure when income per capita reaches $\$2,300$ is the opposite of what RBS find. Diagnostic checks from the Bayesian MF Weibull models applied to the RBS (2007) data show that the obtained Markov chains are stable, have good mixing, and have successfully converged to the desired posterior.

4 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 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 misclassified failure (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 accurate estimates of parameter effects when observed event-failures include cases that in actuality "live on" past their observed-failure point.

We also define this model's conditional posterior distribution and present a slice sampling estimation algorithm (i.e., MCMC method) that allows researchers to conduct Bayesian inference with our model. Here, we provide a dedicated R package for estimating this Bayesian MF survival model as a complement to this paper. Results from extensive Monte Carlo 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 efficiency and have substantially lower RMSEs compared to estimates from regular survival models. Our MF duration model 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 misclassified failure. 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 well help to inform researchers of problematic coding and data collection decisions with respect to failure misclassification. In other cases, these insights and the substantive effects derived from the MF duration model may reveal the theoretical mechanisms that cause political actors to overstate survival-failure in some circumstances but not others.

Notwithstanding these benefits, the model presented can here be extended in four main directions. First, our statistical framework can be used to develop the semi parametric Cox Proportional Hazard [PH] MF model. Although scholars in Comparative Politics and International Relations commonly use parametric survival models such as the Weibull model that we focus on here, Political Scientists also use the Cox PH model for survival analysis. It is plausible that our parametric MF duration model as well as the techniques developed for estimating this 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., Cioffi-Revilla and Landman 1999; Cress, McPherson and Rotolo 1997; Box-Steffensmeier, Radcliffe and Bartels 2005; Linos 2007) 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 these additional duration outcomes from the American Politics, long-range historical, or institutional compliance literatures. Third, we can also note that left-censored survival data is a widespread problem in Comparative Politics and International Relations (Carter and Signorino 2013). In light of this, our MF survival model could also be refined to allow researchers to account for this problem, by estimating the MF model backwards in time, rather than forwards. Finally, the estimator presented in this paper can be further extended to develop a statistical model that econometrically evaluates how spatial factors (e.g., geographic distance or spatial diffusion of the main outcome variable) can simultaneously affect both the hazard of the event of interest and probability of misclassified failure or more generally, the probability that the population of interest emerges from two distinct data generation processes. To this end, scholars have been developing a new class of Bayesian Mixture Cure models that allow (i) spatial correlation in the survival stage of the cure model (e.g. Banerjee, Carlin and Gelfand 2014) or (ii) spatial correlations (by including spatial frailty) in both the survival and split population survival stage of the cure model (Joo and Mukherjee 2018a,b).

References

Balcells, Laia and Stathis N Kalyvas. 2014. "Does warfare matter? Severity, duration, and outcomes of civil wars." *Journal of Conflict Resolution* 58(8):1390–1418.

Balch-Lindsay, Dylan and Andrew J Enterline. 2000. "Killing time: The world politics of civil war duration, 1820–1992." *International Studies Quarterly* 44(4):615–642.

Banerjee, Sudipto, Bradley P Carlin and Alan E Gelfand. 2014. *Hierarchical modeling and analysis for spatial data*. Crc Press.

Beger, Andreas, Cassy L. Dorff and Michael D. Ward. 2014. "Ensemble Forecasting of Irregular Leadership Change." *Research & Politics* pp. 1–14.

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

Beger, Andreas, Daniel W. Hill, Nils. W. Metternich, Shahryar Minhas and Michael 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, Janet M. and Chrisopher Zorn. 1999. "Modeling Heterogeneity in Duration Models." Paper Presented at the 1999 Summer Meeting of The Political Methodology Society.

Box-Steffensmeier, Janet M., Peter M. Radcliffe and Brandon 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, Janet M., Suzanna De Boef and Kyle Joyce. 2007. "Event Dependent and Heterogeneity in Duration Models: the Conditional Frailty Model." *Political Analysis* 15(3):237–256.

Brandt, Patrick T., T. David Mason, Mehmet Gurses, Nicolai Petrovsky and Dagmar 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, Halvard and Scott Gates. 2002. "The geography of civil war." *Journal of Peace Research* 39(4):417–433.

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

Carter, David B and Curtis 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, Claudio and David 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, Claudio and Todd Landman. 1999. "Evolution of Maya Polities in the Ancient Mesoamerican System." *International Studies Quarterly* 43(4):559–598.

Coghlan, Benjamin, Pascal Ngoy, Flavien Mulumba, Colleen Hardy, Valerie Nkamgang Bemo, Tony Stewart, Jennifer Lewis and Richard J Brennan. 2009. "Update on mortality in the Democratic Republic of Congo: results from a third nationwide survey." *Disaster medicine and public health preparedness* 3(2):88–96.

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

Collier, Paul, Anke Hoeffler and Måns Söderbom. 2004. "On the Duration of Civil War." *Journal of Peace Research* 41(3):253–273.

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

Cunningham, David E. 2006. "Veto players and civil war duration." *American Journal of Political Science* 50(4):875–892.

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

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

Joo, Minnie M., Bomin Kim, Benjamin E. Bagozzi and Bumba Mukherjee. 2018. *BayesMFSurv:* An R Package for Bayesian Split Population Survival Models for Duration Data With Misclassified Failure. R package version 0.1.0.

Joo, Minnie M. and Bumba Mukherjee. 2018*a*. "Bayesian Cure Models with Spatial Frailties." Working Paper, Department of Political Science, Penn State University.

Joo, Minnie M. and Bumba Mukherjee. 2018*b*. *BayesSpatialCure:* An R Package for Bayesian Spatial Cure Models for Duration Data with Split Population and Spatial Frailties. R package version 0.1.0.

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

Kreutz, Joakim. 2010. "How and when armed conflicts end: Introducing the UCDP Conflict Termination dataset." *Journal of Peace Research* 47(2):243–250.

Larmer, Miles, Ann Laudati and John F Clark. 2013. "Neither war nor peace in the Democratic Republic of Congo (DRC): profiting and coping amid violence and disorder." 40:1–12.

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

Linos, Katerina. 2007. "How Can International Organizations Shape National Welfare States? Evidence From Compliance With European Union Directives." *Comparative Political Studies* 40(5):547–570.

Lischer, Sarah Kenyon. 2015. *Dangerous sanctuaries: refugee camps, civil war, and the dilemmas of humanitarian aid*. Cornell University Press.

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

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

Reenock, Christopher, Michael Bernhard and David Sobek. 2007. "Regressive socioeconomic distribution and democratic survival." *International Studies Quarterly* 51(3):677–699.

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

Shimray, Ungshungmi A. 2001. "Ethnicity and socio-political assertion: the Manipur experience." *Economic and Political Weekly* 36(39):3674–3677.

Straus, Scott. 2012. "Wars do end! Changing patterns of political violence in sub-Saharan Africa." *African Affairs* 111(443):179–201.

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

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

Thyne, Clayton 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, Joseph K. and Laura Dugan. 2014. "Survival of the Fittest: Why Terrorist Groups Endure." *Perspectives on Terrorism* 8(2):2–23.