Discrete Measurement of Time and Interval Censoring in Event History Analysis

Bruce A. Desmarais
University of North Carolina at Chapel Hill
bdesmara@email.unc.edu

March 6, 2010

Abstract

A problem in event history analysis is that time is measured imprecisely. Events are *typically* known to occur within discrete time units (e.g. day, month or year). Discrete measurement of the start and end time of an event leads to a known interval within which the event duration falls. The event duration is *interval censored*. When ignored, interval censoring is shown to introduce considerable bias to parameter estimates and heighten the risk of inference errors. I show that treating the duration as an interval reduces bias and improves the performance of hypothesis tests. Replications of analyses from four political science articles in leading journals demonstrate that substantive inferences depend on the use of appropriate methods for interval censored duration data. I also develop a software package that can be used to estimate the Cox proportional hazards model with interval censoring.

Note: *Simulation code and replication data will be made available online if this manuscript is accepted for publication. I would like to thank Tom Carsey, Jim Stimson, Skyler Cranmer, Jeff Harden and Georg Vanberg for assistance in this project. I would also like to thank the authors of the analyses I replicate for making their replication materials available.*

1 Measurement and the Timing of Political Events

Time-to-event models – known also as survival, event history, duration and reliability models – have been used to test theory spanning many areas of political science . For examples see Box-Steffensmeier, Arnold and Zorn (1997) and Boehmke (2006) – applications to position taking in the U.S. Congress, Vining, Zorn and Smelcer (2006) – a study of judicial careers, and Goertz, Jones and Diehl (2005) – a work on international rivalry. The timing of an event is never measured precisely. An event is known to occur within a discrete time unit (e.g. day, month or year). Discrete measurement of continuous time leads to an interval within which an event duration is known to fall. The appropriate form of the dependent variable in a time-to-event model is therefore an interval and not a point. The duration is *interval censored*. The common approach in political science, treatment of durations as if they were measured precisely, constitutes an overstatement of certainty and results in downward-biased confidence intervals and – through misspecification – biased parameter estimates.

It is true that imprecise measurement is ubiquitous in data analysis, but discrete measurement of time presents the researcher with much more information regarding the form of the measurement error than do other imprecise measurement mechanisms. Since the discrete interval within which the duration falls is known, so are the upper and lower bounds of that interval. The use of estimators designed for interval outcome data mitigate the problems introduced by treating time as precisely measured. Due to the impossibility of precise measurement of continuous time, *all* duration data is interval censored. Technically speaking – interval censored estimation should be used in any analysis of time-to-event data.

In practice it is important to understand the costs and benefits of implementing the interval censored method. In the current study I describe precisely how discrete measurement censors time and how this censoring can be accounted for in the analysis of time as a dependent variable. Then I present a Monte Carlo study in which I demonstrate the superiority of an algorithm proposed by Pan (2001) for estimation of the Cox proportional hazards model (Cox model for short) under interval censoring. Previously unavailable in any public software release, I have programmed the algorithm of Pan (2001) in the statistical software `R` as `MICox`. Lastly, I re-estimate event history models from four studies published in three leading political science journals. In the replication study I show that a number of past substantive inferences change under interval censored estimation.

2 Discrete Measurement and Interval Censoring

2.1 Illustration of the Problem

When studying the duration of some event, it it standard to measure the duration by subtracting the start time (t_s) from the end time (t_e). These start and end times are approximated by a discrete measurement process. That is, the researcher has some measurement tool, a calendar or stopwatch for instance, and the time that the tool reads when the event starts and ends are the values assigned to (t_s) and (t_e) respectively. If the precision of the measurement tool is annual, then all events that start between January 1, 2000 and December 31, 2000, and end between January 1, 2001 and December 31, 2001 will be assigned a duration of one year (2001-2000 =1). Note that the duration that leads to this measurement of one year could actually be as short as one day (12/31/2000-1/1/2001) or as long as one day short of two years (1/1/2000-12/31/2001). The problem arising from exact treatment of interval censored time is that both of these events would be treated as exactly equal in duration, rather than between one and two years.

To illustrate the scope of inferential errors that can arise due to discrete measurement of time, here I discuss a synthetic example of ordering the durations of two events. If two durations (A and B) are measured to be equal or adjacent, it is possible that $A > B$, $B > A$, or $B = A$. It was demonstrated above that events of different length can be falsely assigned the same duration. The following is an example of how the order of durations can actually be opposite of that measured. Assume A is a conflict that begins in 1991 and ends in 1992, thus receiving a duration of one, and that B is a conflict that starts in 1991 and ends in 1993, receiving a duration of two. If A starts on 1/1/1991 and ends on 12/31/1992, its daily duration is 730 days, and if B starts on 12/31/1991 and ends on 1/1/1993, it has a daily duration of 367 days. In annual terms, A is only one half the duration of B, but in daily terms, A is nearly twice as long as B. This is an example of the most extreme distortion that can be introduced via discrete measurement.

2.2 From Point to Interval

More generally, the discrete measurement of the duration of an event leads to a known interval within which the actual duration falls [1]. This interval arises from the two intervals indicated by the discrete measurement of t_s and t_e. If an event is measured to start at time t_s, its exact start time is between t_s and $t_s + 1$. If it is measured to end at time t_e, the actual end time t_e is known to be between t_e and t_e+1. These two intervals are used to construct the known upper and lower bounds within which the exact duration lies. The longest possible duration, given t_s and t_e is given by the difference between the latest possible end time and the earliest possible start time $(t_e + 1 - t_s)$ and the shortest possible duration is given by the

[1] In all of the examples and derivations it is assumed that the precision of measurement has been scaled to the integer level (e.g. if the precision is daily and event that lasts one day is of duration one, not 1/365 years. This assumption retains the generality of results and reduces the amount of jargon in the discussion.

difference between the earliest end time and the latest start time $(t_e - t_s - 1)$. This gives an interval within which the exact duration lies. The interval for the true duration (d) is given by:

$$t_e - t_s - 1 \leq D \leq t_e - t_s + 1 \tag{1}$$

The interval given in 1 contains the points within ± 1 of the measured duration $m = t_e - t_s$. The fundamental step in accounting for discrete measurement of time in event history analysis is to adapt the estimator to the fact that the appropriate dependent variable is the interval $[m-1, m+1]$ and not the point m.

2.3 Interval Censored Estimation

The general approach to the analysis of discrete measured continuous-time durations advocated here is to treat the dependent variable as interval censored on the interval $[m-1, m+1]$.[2] To motivate the derivation of interval censored estimators, recognize that the probability (possibly conditional upon covariates) of observing measured duration m is equal to the probability that the true duration $d \in [m-1, m+1]$. The likelihood ($L(\cdot)$) of a sample of discrete measured (and thus interval censored) durations conditional on independent variables X is

[2] Many authors will note that the form of censoring introduced by discrete measurement is a form of double-censoring (e.g. both the start and end times are known only up to a defined interval (Jewell, Malani and Vittinghoff 1994; Pan 2001; Cai and Cheng 2004). As demonstrated above, discrete measurement leads to an interval within which the exact duration is known to lie, and thus the interval censoring induced by discrete measurement is a special case of double-censoring

$$L(\boldsymbol{m}, \boldsymbol{X}, \boldsymbol{\theta}) = \prod_{i=1}^{n} \left(\int_{m_i - 1}^{m_i + 1} f(z|\boldsymbol{x}_i, \boldsymbol{\theta}) dz \right)^{\delta_i} S(m_i|\boldsymbol{x}_i, \theta)^{1-\delta_i} \quad (2)$$

$$L(\boldsymbol{M}, \boldsymbol{X}, \boldsymbol{\theta}) = \prod_{i=1}^{n} \left[F(m_i + 1|\boldsymbol{x}_i, \boldsymbol{\theta}) - F(M_i - 1|\boldsymbol{x}_i, \boldsymbol{\theta}) \right]^{\delta_i} S(m_i|\boldsymbol{x}_i, \boldsymbol{\theta})^{1-\delta_i}. \quad (3)$$

Where $F(\cdot)$ is the cumulative distribution function, $S(\cdot)$ is the survival function (i.e. $1 - F(\cdot)$) and δ_i is a binary indicator of failure. The exact form of the likelihood function depends on the choice of $f(\cdot)$. The likelihood in equation 2 can be compared to that appropriate for precisely measured data, which is equal to

$$L(\boldsymbol{m}, \boldsymbol{X}, \boldsymbol{\theta}) = \prod_{i=1}^{n} f(m_i|\boldsymbol{x}_i, \boldsymbol{\theta})^{\delta_i} S(m_i|\boldsymbol{x}_i, \boldsymbol{\theta})^{1-\delta_i}. \quad (4)$$

The only difference between equations 4 and 2 is that there is an integral over all of the plausible durations for an observation that fails in equation 2. Note that the likelihood contribution of an observation that does not fail is the same in each equation - $S(m|\boldsymbol{x}, \boldsymbol{\theta})$. Recall that a duration measured to fail at the m^{th} unit of measurement is known to be in the interval $[m - 1, m + 1]$. If an observation does not fail by the m^{th} period, the next period is the first period in which a failure could be recorded, in which case the true event time would be in $[m, m + 2]$. Thus, if all that is observed about an event is that it does not occur by the m^{th} point, all that can be said is that it has a duration of at least m, and is right censored at m (Box-Steffensmeier and Jones 1997).

All of the familiar distributions such as the Weibull, log-normal and log-logistic are readily implemented within this framework (Odell, Anderson and D'Agostino 1992; Kim 1997; Betensky, Rabinowitz and Tsiatis 2001). Since the likelihood function is correctly specified if discrete measured duration data is treated as interval censored, the regression and duration-dependence parameters that are estimated represent the minimum-variance asymptotically unbiased estimates of the parameters of the continuous-time distribution.

Estimation is not as straightforward in the case of the Cox Proportional Hazards Model. The likelihood function is not fully specified in the Cox model. The identifying assumption in the model is that the hazard rate is proportional to $e^{\boldsymbol{\beta}'\boldsymbol{x}}$ (Box-Steffensmeier and Jones 1997). Since the exact form of $F(\cdot)$ is not specified in the Cox model, and technically only the order of the sampled events enters the likelihood function rather than the actual magnitudes, complicated algorithms have been derived to deal with uncertainty regarding the event ordering (Satten, Datta and Williamson 1998; Pan 1999, 2000, 2001; Pan and Chappell 2002; Kim 2003). In either parametric or the semi-parametric Cox model, parameter estimates and standard errors are biased if the midpoint of the interval $(t_e - t_s)$ is imputed as the observed failure time, and the estimator used does not account for interval censoring.

3 Midpoint Imputation Bias

Inference on many components of survival data - including non-parametric estimation of the survival curve, semi-parametric estimation of the regression function, and parametric estimation of the full conditional distribution - have been shown to be biased under *single-imputation* (i.e. the use of the measured duration as exact). The most common form of imputation in applications is midpoint imputation (i.e. $t_e - t_s$) (Goggins, Finkelstein, Schoenfeld and Zaslavsky 1998). The degree of this bias has been the focus of much inquiry.

Since midpoint imputation results in a misspecified likelihood function (i.e. equation 4 is used when the interval censored likelihood - appearing in equation 2 - is appropriate), there is good reason to expect that the inferential properties of maximum likelihood do not apply to estimation with midpoint imputed durations. It is well established that estimates are often biased under maximum likelihood estimation with a misspecified likelihood function (White 1982; Heagerty and Kurland 2001). To examine the implications of midpoint imputation in specific cases, the common strategy employed in the literature has been Monte Carlo experimentation.

A variety of simulation experiments have been designed to asses the degree of bias induced by exact-time estimation on midpoint imputed data. In one of the first works to propose an alternative to midpoint imputation for the Cox model with interval censored data, with a simulation study involving 1000 replications and exponentially distributed failure times dependent on a single covariate, Goggins et al. (1998) show that there is an approximate 10% downward bias in the estimate of the regression parameter and that the estimated asymptotic standard error is 6% less than that corresponding to their interval censored estimator. Satten (1996) finds a consistent 40-50% downward (in magnitude) bias in Cox regression estimates under midpoint imputation in a simulation study with 100 replications and data generated from a Weibull distribution with increasing hazards. In a simulation study of a Weibull accelerated failure time model, Odell, Anderson and D'Agostino (1992) find that the asymptotic 95% confidence intervals for the midpoint imputed estimator exhibit empirical coverage probabilities around 85% and under certain conditions - decreasing hazards (i.e. negative duration dependence) and large sample size (1,200) - the coverage probability for a 95% CI can be as low as 17%. In their Monte Carlo study - with two covariates, 500 replications and Weibull distributed interval censored failure times with increasing hazards - Wei, Ying, Chaloner and Stapleton (2009) also find that coverage probabilities in the Cox model are biased downward, but only by 1-2%.

Overall, the findings on the exact degree of bias have been mixed. Yet, a couple of findings are consistent across the Monte Carlo studies. First, exact-time estimation on midpoint imputed data poses the risk of type-I inference errors. Confidence intervals are biased downwards. This is evident in the numerous studies that find downward-bias in the coverage probabilities. The intuition underlying this result can be linked to the work on multiple imputation (Rubin 1996; King, Honaker, Joseph and Scheve 2001). Multiple candidates are generated for each missing datum to represent uncertainty in the imputed values. This uncertainty is then translated into the estimates computed on the multiple datasets. Imputing a midpoint is a single, certain imputation. In this sense, midpoint imputation imposes too much certainty on the exact value of the failure times. This biased degree of certainty translates into downward biased confidence intervals for the estimates.

The second finding about midpoint imputation is that it creates downward bias in the magnitude of the regression estimates. The intuition underlying this result can be motivated from the perspective of measurement error. Meier, Richardson and Hughes (2003) show that the estimates of effects in the Cox model are biased toward zero when measurement error is introduced to the survival times, even when the measurement error is unbiased (i.e. when the expected measurement is equal to the actual duration). Imputing the midpoint for an observed duration constitutes a measurement error insofar as the true duration is not exactly equal to $t_e - t_s$. These findings serve as strong motivation to implement proper interval censored estimation of discretely measured event durations in political science. As noted below, techniques for parametric regression with interval censored data are well established and available in many popular software packages. This is not the case for implementations of the Cox model. In the next section I compare the Cox model with midpoint imputed data with an original implementation of the method appearing in Pan (2001).

4 The Cox Proportional Hazards Model for Interval Censored Data: A Monte Carlo Study

The estimation of parametric survival models for interval censored data is performed with a straight-forward application of numerically maximizing the likelihood function, which takes the form of equation 2. These procedures are implemented in the `STATA` (StataCorp

2007) add-on `INTCENS` (Griffin 2005), and in the R (R Development Core Team 2008) package `survival` (Therneau and original R port by Thomas Lumley 2009). Software for the analysis of interval censored data with the Cox Proportional Hazards model is not nearly as developed as that for fully parametric analysis. Currently, one of the many algorithms proposed to analyze interval censored data with Cox regression has been implemented in a major software package - R . Gomez, Calle, Oller and Langohr (2009) provide an extensive review of the `intcox` package (Henschel, Heiss and Mansmann 2009), complete with R code and example data available on the web, which implements the technique proposed by Pan (1999). A major shortcoming of this software - and the reason why I do not consider it beyond this point - is that it can *only* handle interval censored data, and cannot accommodate right censored observations. The method that I use is that proposed by Pan (2001). This is a multiple imputation based approach and I call it `MIcox`.[3]

Pan (2001) shows that the algorithm proposed by Sun, Liao and Pagano (1999), which involves the computationally intensive deterministic solution of a complex set of estimating equations, can be implemented with a multiple-imputation approach (Rubin 1987). Moreover, `MIcox` requires only the standard estimation routines for the Cox model with exactly-measured and right censored durations. Noted by Pan (2001), another advantage of this routine over the estimation procedure proposed by Sun, Liao and Pagano (1999) and the other interval censored methods for the Cox model is that the usual diagnostics, such as tests for proportional hazards (Box-Steffensmeier and Zorn 2001), can be utilized by averaging over the imputations.

The first step in `MIcox` is to estimate the marginal (i.e. univariate) density - $\hat{H}$ - of $\boldsymbol{d}$ -

[3] I have programmed the algorithm used for estimation with `MIcox` in R , and am currently developing a package for public distribution.

the sample of durations - with a non-parametric routine adapted for interval censored data such as that proposed by Turnbull (1976) or (Wellner and Zhan 1997).[4] `MIcox` proceeds by generating k imputed datasets, where a new exact duration is drawn from $\hat{H}$ for each interval censored observation, conditional on the i^{th} imputed observation being in the interval $[m_i - 1, m_i + 1]$. This leads to k datasets composed of only exact and right censored observations. Standard Cox regression is then performed on each dataset, producing k estimates of the regression coefficients - $\boldsymbol{\beta} := (\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)} .. \boldsymbol{\beta}^{(k)})$ - and k estimates of the variance of $\boldsymbol{\beta}$ - $\boldsymbol{\sigma} := (\boldsymbol{\sigma}^{(1)}, \boldsymbol{\sigma}^{(2)} .. \boldsymbol{\sigma}^{(k)})$. These results are then combined in a manner that accounts for the imputation-to-imputation variability of the results for $\boldsymbol{\beta}$ – forming the variance estimate as the cross-imputation variance in the estimate of $\boldsymbol{\beta}$ added to the average within-imputation variance (Rubin 1987; King et al. 2001). The estimate of the regression coefficients - $\hat{\boldsymbol{\beta}}$ - is given by

$$\hat{\boldsymbol{\beta}} = \frac{1}{k} \sum_{i=1}^{k} \boldsymbol{\beta}^{(i)}. \tag{5}$$

And the estimate of the variance of $\hat{\boldsymbol{\beta}}$ is

$$\hat{\boldsymbol{\sigma}^{\mathbf{2}}} = \frac{1}{k} \sum_{i=1}^{k} \boldsymbol{\sigma}^{(i)} + \left(1 + \frac{1}{k}\right) \frac{\sum_{i=1}^{k} (\boldsymbol{\beta}^{(i)} - \hat{\boldsymbol{\beta}})(\boldsymbol{\beta}^{(i)} - \hat{\boldsymbol{\beta}})}{k-1}. \tag{6}$$

It can be seen from the fraction $\frac{1}{k}$ in the second term of equation 6 that efficiency is improved with additional imputations, but there is a rapidly decreasing return for additional imputations beyond ten or so. It is standard in the literature to use twenty or less imputations (Rubin 1996; King et al. 2001).

[4]For `MIcox`, I use the algorithm proposed by Wellner and Zhan (1997), which is available in the `R` package `Icens` (Gentleman and Vandal 2009), to estimate $\hat{H}$.

4.1 Simulation Design

The objective of the Monte Carlo study is to examine the performance of `MICox` relative to midpoint imputation under standard interval censoring conditions *and* to determine whether `MICox` exhibits any performance short-falls when censoring is negligible. [5] The durations are generated from a Weibull distribution, conditional on a single standard-normally distributed covariate. The covariate enters multiplicatively to the hazard rate in the form given in equation [24] of Box-Steffensmeier and Jones (1997). The regression coefficient is set at 1.0 for all simulations. The simulation is run with sample sizes (N) of 25, 50, 100, 200 and 400. The mean of the baseline duration (i.e. the duration when the covariate is zero) is set at 10, 15, 25, 45 and 85. The duration dependence parameter (α) is set at 0.75 and 2.0 to examine performance under increasing and decreasing hazards (e.g. negative and positive duration dependence) respectively. The precision of measurement is set at the integer level. Varying the mean varies the amount of information removed from the sample through interval censoring. The ordering of the durations, which fully determines the estimates in the Cox model, is much more likely to be eschewed by integer interval censoring in a sample with a mean of 10 than in one with a mean of 85. For each combination of sample size, mean duration and duration dependence, I run 5000 replications. The multiply imputed estimates

[5] As it is beyond the scope of the current paper, I do not attempt to render a comprehensive assessment of the every-day conditions under which the `MICox` algorithm outperforms midpoint imputation. Pan (2001) and (Sun, Liao and Pagano 1999) have already theoretically, and through their own Monte Carlo studies, established the superiority of the interval censored approach to midpoint imputation in terms of bias and efficiency. My goal here is to determine whether my particular implementation of `MICox` in `R` outperforms midpoint imputation - as one would predict from reading Pan (2001) - under reasonable conditions.

for `MICox` are based on 20 random samples from $\hat{H}$.[6]

The censoring process used in the Monte Carlo study should reflect that which is induced by discrete measurement of time. If the exact duration is $t+\delta,\ 0 \leq \delta < 1$, then it is possible it was measured to last either t or $t+1$ measurement periods. Recall that a duration that is measured to last m periods can have an exact value between $m-1$ and $m+1$. Thus, if $m = 4$, it is known that the exact duration is between 3 and 5, and if $m = 5$, the exact duration is between 4 and 6. Inverting this logic, an exact duration known to be in the interval $[4, 5)$ - because it is simulated - could have been measured as $m = 4$ or $m = 5$. To censor the durations generated in the simulation I randomly assign them to have been measured as either t or $t+1$. Moreover, the probability of choosing the measurement $t+1$ is equal to δ, making it morel likely, for instance, that a 4.9 will be measured as a 5 than a 4. This selection maps all continuous values onto plausible two-unit measurement intervals.

4.2 Simulation Results

The results of the Monte Carlo studies are presented in figures 1-3. Similar to the findings of Odell, Anderson and D'Agostino (1992), the benefits from using proper interval censored estimation relative to midpoint imputation depend heavily on the duration dependence parameter.[7] Overall, when duration dependence is negative (e.g. increasing hazards), the improvement in estimator performance relative to midpoint imputation is much more

[6]The Breslow method for handling ties (Box-Steffensmeier and Jones 2004), which is the method used in all of the original studies I replicate in the next section, is used to handle ties in the exact Cox estimation.

[7]To my knowledge, this conditionality on duration dependence has not been thoroughly studied, and is therefore an open topic for inquiry.

apparent than when duration dependence is positive.[8] In order to judge the performance of midpoint imputation and `MICox`, I compute the bias in the estimate of the regression coefficient, mean absolute error and 95% confidence interval coverage probability for each combination of sample size, duration dependence parameter, and mean baseline duration.[9]

The bias of the estimates is presented in figure 1. The most relevant pattern in terms of the current study is that the use of `MICox` reduces the bias by approximately 0.15 across all simulation parameters when there is negative duration dependence. Examining the scale of the y-axes in figure 1 reveals that the degree of the bias is an order of magnitude larger when there is negative duration dependence ($\alpha = 2$). The bias ranges between -0.40 and 0.013 with negative duration dependence and between -0.04 and 0.06 with positive duration dependence - a range that is possibly completely attributable to simulation error and not actual bias. There is no clear relationship between the bias and the sample size. When less information is removed from the data through censoring, and the average baseline duration is higher, the magnitude of the bias is generally lower.

[8] Both of the studies considered for replication in the current paper that utilize parametric regression models - Fearon (2004) and Hartzell and Hoddie (2003) - find their data exhibits negative duration dependence in a Weibull specification, with scale parameter estimates of 1.21 and 1.18 respectively.

[9] Denote the i^{th} estimates of the regression parameter and standard error $\hat{\beta}_i$ and $\hat{\sigma}_i$ respectively. The bias is computed as $\frac{1}{500}\sum_{i=1}^{500}\hat{\beta}_i - 1$, the mean absolute error as $\frac{1}{500}\sum_{i=1}^{500}|\hat{\beta}_i - 1|$, and the coverage probability as $\frac{1}{500}\sum_{i=1}^{500} I\left(\hat{\beta}_i - 1.96\hat{\sigma}_i \leq 1 \leq \hat{\beta}_i + 1.96\hat{\sigma}_i\right)$, where $I(\cdot)$ is the indicator function which is equal to 1 (0) if its argument evaluates to true (false).

[Insert Figure 1 here]

The mean absolute error of the estimates is presented in figure 2. As with the bias, in terms of mean absolute error, there is a clear advantage of $30-50\%$ in using `MICox` under negative duration dependence. There is no clear advantage when duration dependence is positive. As sample size increases, the lower realized bound on mean absolute error decreases. Also, when duration dependence is negative, there is a clear negative relationship between the average baseline duration and mean absolute error.

[Insert Figure 2 here]

The estimated coverage probabilities of the 95% confidence intervals from the Monte Carlo study are presented in figure 3. In a continuation of the theme of findings, there are major gains to be had from using `MICox` when there is negative duration dependence. Also under negative duration dependence, I find that (1) any notable bias in the coverage probability is downward, (2) the bias decreases as the mean baseline duration increases, and (3) the bias worsens as sample size increases. When there is positive duration dependence, the estimated coverage probabilities range between 91 and 98%, but there is no clear advantage of `MICox`. Nor is there a clear relationship between coverage probability and any of the other simulation parameters when there is positive duration dependence.

[Insert Figure 3 here]

The finding in the Monte Carlo study is that, under negative duration dependence, `MICox` outperforms midpoint imputation in terms of bias, mean absolute error and coverage probability. The results don't point in the direction of either estimator when duration dependence is positive. It may be tempting to advise analysts to attempt to assess the duration dependence in their data prior to choosing an estimator, where there would be little consequence of midpoint imputed estimation when duration dependence is positive, but this would be an unwise recommendation. A parametric routine is necessary to assess the shape of the baseline hazard, and the justification for using the Cox model is that it is not necessary to estimate the baseline hazard. Thus, if the analyst deems the Cox model appropriate, and the parametric family of the baseline hazard is unknown, *a priori* assessment of duration dependence is not possible without an unwarranted assumption. Since it is uniformly better to use `MICox` under negative duration dependence and there does not seem to a disadvantage under positive duration dependence, I recommend that analysts utilize `MICox` in all situations. In the next section, I demonstrate the implications of proper interval censored estimation with replication of the analyses in some major studies that use duration data.

5 Implications of interval Censoring in Political Science

In this section, I replicate models that appear in recent applications from legislative studies (Box-Steffensmeier, Arnold and Zorn 1997), judicial politics (Shipan and Shannon 2003) and the study of civil war (Hartzell and Hoddie 2003; Fearon 2004) - three areas that see frequent use of time-to-event data.[10] I replicate two analyses that use the Cox

[10]For contemporary examples of analysis of time-to-event data in these sub-fields see Spriggs and Hansford (2001); Bell (2002); Benesh and Reddick (2002); Binder and Maltzman (2002); Strom and Swindle (2002); Peterson, Grossback, Stimson and Gangl (2003); Senese and Quackenbush (2003); Fortna (2004); Caldeira and Zorn (2004); Cunningham (2006) and Krustev (2006) among others.

model exclusively, one that uses both Cox and Weibull models, and another that uses the Weibull only. These examples, along with all previous implementations of duration analysis in political science, inappropriately assume that time is measured exactly or that time itself is discrete. The objective of this replication study is to examine whether the validity of inferences based on data that is commonly studied in political science can be questioned on the basis of these inappropriate assumptions. Due to the availability of software to implement interval censored techniques, a consistent finding of even a moderate change in inferences provides a strong motivation to adopt interval censored procedures in estimation.

The replication results are presented in a standardized manner in each study. A single table for each replication contains all of the original and interval censored models. For each regression model, a column giving the midpoint imputed estimate is presented first, then one giving the interval censored estimate, and a third presenting the percentage change from the midpoint imputed to the interval censored estimate (%Δ). This third column can be used to assess the effect of accommodating the discrete measurement of time relative to imputing the midpoint. If the use of interval censored estimation changes the result of a hypothesis significance test performed at the 0.05 level (one-tailed), then the value of %Δ is reported as Δ^{+}. The original results in each example were replicated exactly.[11]

5.1 Delay in Senate Confirmation of Supreme Court Nominees

Shipan and Shannon (2003) study the time it takes the U.S. Senate to confirm nominees to the U.S. Supreme Court. Utilizing a theoretical spatial model of the confirmation process, they formulate a prediction that the confirmation process will take longer when the Senate

[11] All of the data for the replications was acquired from the authors of the original studies via either email request or downloaded from their website.

and President are ideologically distant from each other. They test this hypothesis with a Cox regression model of the "days in session" it takes to confirm a justice for all Supreme Court nominees from the Civil War to 2002. The average duration of confirmation is 39.09 days. They use two highly correlated variables, *President-Majority Party Distance* and *Divided Control* to test their primary hypothesis. Due to the high correlation they specify two equations, leaving one out in each of them. Replications of both of their models appear in table 1.

[Insert Table 1 here]

The primary independent variables of interest in (Shipan and Shannon 2003) - *Divided Control* and *President-Majority Party Distance* - have a statistically significant effect in the expected direction on the duration of confirmation in both the midpoint imputed and interval censored models. In further confirmation of the theoretical argument, the primary inference in the article - that ideological distance between the Senate and the President lengthens confirmation proceedings - is unaffected by the choice of estimator.

This is not true for all of the independent variables in the model. First of all, the magnitude of many of the estimated effects and z-statistics differ by $20 - 40\%$ from the midpoint imputed to interval censored estimates. Additionally, 0.05-level hypothesis tests are changed for three of the variables in the model. In both models *Federal District Court Experience* has a statistically significant positive effect on the hazard of confirmation proceedings in both of the midpoint imputed models, but is not significant in either of the interval censored estimates. In the replication of model [4], the interaction between *Critical Nomination* and *Divided Control* is not significant with midpoint imputed estimation, but has a significant positive effect on the hazard with `MICox`. Lastly, in model [3] the duration of confirmation under an *unelected president* is not significantly different with midpoint imputed estimation, but with `MICox` the hazard is higher under an *unelected president*. In short, many of the statistical and substantive inferences rendered from the analysis in Shipan and Shannon (2003) depend on the use of `MICox`.

5.2 The Timing of Position Taking in Congress

In one of the seminal applications of survival analysis to political durations, Box-Steffensmeier, Arnold and Zorn (1997) study the strategic timing of position-taking in the U.S. House of Representatives. Specifically, they model the time it takes House members to take a position on the *North American Free Trade Agreement*. The central hypothesis they test is that members will take a position quickly when relevant publics - such as interest groups, constituents and policy leaders - send a clear preferential signal on the legislation. They collect data on the entire U.S. House of Representatives in the 103^{rd} Congress. The dependent variable is the number of days after August 11, 1992 until the member took a position. The average duration is 403 days. They include a number of relevant *Constituency, Interest Group, Institutional* and *Individual* factors as independent variable in the model. The Cox model is used to analyze the data. The results of the replication appear in table 2.

[Insert Table 2 here]

The substantive argument made by Box-Steffensmeier, Arnold and Zorn (1997) garners marginally stronger empirical support with the use of `MICox` relative to midpoint imputation. One of the variables they use to test the central contention of the article - that members take a position faster when they receive clearer signals from constituents - is tested with the district-level percentage of votes for Perot in the 1992 presidential election. Because Perot made the support for the defeat of NAFTA a central issue in his campaign, Box-Steffensmeier, Arnold and Zorn (1997) expect that members representing districts exhibiting notably low or high - but not moderate - levels of support for Perot will take a position earlier.

If members *noisily* equated the proportion supporting Perot with the proportion in their district objecting to NAFTA, there were no observed instances of strong objection to NAFTA in the data. The maximum vote for Perot was 33%. There were, however, many instances of low support. Approximately half the districts voted for Perot at a rate of 25% or less. All of this is illustrated in figure 4. Figure 4 gives the effect of the vote for Perot. It is seen here that - under estimation with `MICox` - the effect of the Perot vote on the hazard is statistically significantly positive for vote levels of approximately $5-20\%$, and no different from zero for levels above 20%. If $5-20\%$ is considered moderate support for Perot, the estimates with `MICox` provide at least partial support for the expectation Box-Steffensmeier, Arnold and Zorn (1997) have for the effect of support for Perot – a hypothesis not supported by the original results. Though none of the other hypothesis tests in this article were dependent upon the use of `MICox`, more than half of the coefficient and z-statistic estimates differed by double-digit percentages from the midpoint imputed to the `MICox` estimates.

[Insert Figure 4 here]

5.3 The Duration of Civil War

Fearon (2004) constructs a game-theoretic explanation of the duration of civil conflict. The model produces four empirical expectations:

1. Wars originating as coups or popular revolutions are shorter than those originating in other manners.

2. Deemed a *sons of the soil* conflict, a civil war will last longer if it is between state-supported migrants and an ethnic minority that inhabits an area rich in natural resources.

3. When the rebels have a contraband-based stream of income, the civil war will last longer.

4. Anti-colonial wars - those between non-contiguous parties - are shorter.

As "inductive" evidence of these patterns, Fearon (2004) presents a Weibull accelerated failure time regression of the annual duration of 128 civil wars originating between 1945 and 1999. The average civil war in the sample lasts for 8.95 years. [12] The primary variables of interest are *Coup/revolution* - an indicator of whether origination was the result of a coup or popular revolution, *Not contiguous* - an indicator of whether the war is colonial, *Sons of the Soil* - an indicator that the war meets the criteria described in hypothesis (4) above, and *Contraband* - an indicator that the rebels earn income from the sale of contraband such as drugs. In the original analysis, all of the variables except *Not contiguous* are statistically significant at the 0.05 level (two-tailed) in the direction that supports the theoretical contentions above. I only replicate the first model presented in table [2] of Fearon (2004).[13] The results are presented in table 3.

[Insert Table 3 here]

The replication of Fearon (2004) demonstrates that some results are robust to the use of either midpoint imputation or interval censored estimation. Unlike the previous replications,

[12]In an accelerated failure time model the covariates are assumed to have a multiplicative effect on the mean of the duration, rather than a multiplicative effect on the hazard rate as in a proportional hazards Weibull or Cox regression.

[13]This is the most parsimonious model, none of the variables added in the others are statistically significant, and both the AIC and BIC favor the first model.

which used the Cox model exclusively, Fearon (2004) uses a Weibull accelerated failure time model. None of the hypothesis tests at the 0.05 level are different between the two models, and the difference in the magnitude of both effects and z-statistics are all single-digit differences. Though this is the case with the duration of civil war - as will be seen next - the same cannot be said for the duration of peace after civil war.

5.4 The Duration of Peace after Civil War

Hartzell and Hoddie (2003) examine how power-sharing institutions embedded in peace agreements can be designed to stabilize the transition from civil war to stable and enduring peace. They model the monthly duration of peace in 38 civil conflicts resolved via the process of negotiations between 1945 and 1998. The average duration of peace in the data is 94.5 months. The primary hypothesis they test is that the duration of peace increases as the number of power-sharing institutions established or empowered in the settlement agreement increases. They construct a composite measure *Power-sharing Institutions* that varies in integer values from zero to four, and is a sum of indicators of whether the agreement contains provisions for political, territorial, military and economic power-sharing institutions. They use both a Weibull proportional hazards model and a Cox model to test their hypothesis. The replication results are presented in table 4.

[Insert Table 4 here]

The original analysis confirmed the primary hypothesis in the study. *Power-sharing Institutions* had a positive and statistically significant effect on the duration of peace. Relative to the Cox model results with midpoint imputation, the magnitude of the effect of *Power-sharing Institutions* decreases by over 60% and is not statistically significant using `MICox`. Also with `MICox`, hypothesis test results at the 0.05 level are different for four of the six control variables in the model. The results of the midpoint imputed and interval censored estimation are more similar in the case of the Weibull estimates. None of the hypothesis test results change in the parametric analysis, but there is reason to be suspicious of the Weibull results. Notably, if the Weibull baseline hazard assumption is appropriate, the interval censored results in the Cox model should not differ greatly from the interval censored parametric results since the Weibull specification is within the proportional hazards framework assumed by the Cox model.

6 Conclusion

discrete measurement introduces biased parameter estimates and downward-biased confidence intervals to duration models when the discrete measured variable is treated as the exact continuous-time value. Limited measurements indicate intervals within which a continuous time variable lies. In short, discrete measurement leads to an interval censored variable. Statistical techniques for optimal parameter estimation under conditions of general interval censoring are already well developed for both the Cox proportional hazard model and parametric estimators. In this study, I demonstrate how duration models can be improved by analyzing the discrete measured duration variable as a systematically interval censored event time. Both bias and the threat of an inference error are reduced when appropriate statistical procedures are used with interval censored data. A replication study in which I re-estimate models from four articles in prominent political science journals demonstrates that parameter estimates and their statistical significance consistently differ when interval censored estimation is used rather than midpoint imputation.

Fortunately, it is not overly burdensome to eliminate uncertainty over whether discrete measurement has distorted estimates through interval censoring. Because interval censored estimation is always appropriate under discrete measurement, analysts can simply estimate their models with midpoint imputed and interval censored estimation techniques and compare the results. If there are major differences between the respective outputs, the analyst should be suspect of the midpoint imputed results. Two major statistical-computing software packages, R and `STATA` have routines developed to implement interval censored estimation. In `STATA` the program `INTCENS` is capable of estimating interval censored parametric duration models. In R the `survival` package can estimate interval censored parametric duration models. Though parametric models have proven useful in many areas of political science, the most popular model in the discipline is the semi-parametric Cox model. A primary contribution of the current project is the development of a package – `MICox` – designed to implement the interval censored Cox proportional hazards model in R .

References

Bell, Lauren Cohen. 2002. "Senatorial Discourtesy: The Senate's Use of Delay to Shape the Federal Judiciary." *Political Research Quarterly* 55(3):589–607.

Benesh, Sara C. and Malia Reddick. 2002. "Overruled: An Event History Analysis of Lower Court Reaction to Supreme Court Alteration of Precedent." *The Journal of Politics* 64(2):534–550.

Berry, Frances Stokes and William D. Berry. 1990. "State Lottery Adoptions as Policy Innovations: An Event History Analysis." *The American Political Science Review* 84(2):395–415.

Betensky, Rebecca A., Daniel Rabinowitz and Anastasios A. Tsiatis. 2001. "Computationally Simple Accelerated Failure Time Regression for Interval Censored Data." *Biometrika* 88(3):703–711.

Binder, Sarah A. and Forrest Maltzman. 2002. "Senatorial Delay in Confirming Federal Judges, 1947-1998." *American Journal of Political Science* 46(1):190–199.

Boehmke, Frederick J. 2006. "The Influence of Unobserved Factors on Position Timing and Content in the NAFTA Vote." *Political Analysis* 14(4):421–438.

Box-Steffensmeier, Janet M. and Bradford S. Jones. 1997. "Time is of the Essence: Event History Models in Political Science." *American Journal of Political Science* 41(4):1414–1461.

Box-Steffensmeier, Janet M. and Bradford S. Jones. 2004. *Event history modeling: A guide for social scientists*. Cambridge University Press.

Box-Steffensmeier, Janet M. and Christopher J. W. Zorn. 2001. "Duration Models and Proportional Hazards in Political Science." *American Journal of Political Science* 45(4):972–988.

Box-Steffensmeier, Janet M., Laura W. Arnold and Christopher J. W. Zorn. 1997. "The Strategic Timing of Position Taking in Congress: A Study of the North American Free Trade Agreement." *The American Political Science Review* 91(2):324–338.

Cai, T. and S. Cheng. 2004. "Semiparametric Regression Analysis for Doubly Censored Data." *Biometrika* 91(2):277–290.

Caldeira, Gregory A. and Christopher Zorn. 2004. "Strategic Timing, Position-Taking, and Impeachment in the House of Representatives." *Political Research Quarterly* 57(4):517–527.

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

Fearon, James D. 2004. "Why Do Some Civil Wars Last so Much Longer than Others?" *Journal of Peace Research* 41(3):275–301.

Fortna, Virginia Page. 2004. "Does Peacekeeping Keep Peace? International Intervention and the Duration of Peace after Civil War." *International Studies Quarterly* 48(2):269–292.

Gentleman, Robert and Alain Vandal. 2009. *Icens: NPMLE for Censored and Truncated Data*. R package version 1.16.0.

Goertz, Gary, Bradford Jones and Paul F. Diehl. 2005. "Maintenance Processes in International Rivalries." *The Journal of Conflict Resolution* 49(5):742–769.

Goggins, William B., Dianne M. Finkelstein, David A. Schoenfeld and Alan M. Zaslavsky. 1998. "A Markov Chain Monte Carlo EM Algorithm for Analyzing Interval-Censored Data under the Cox Proportional Hazards Model." *Biometrics* 54(4):1498–1507.

Gomez, Guadalupe, M. Luz Calle, Ramon Oller and Klaus Langohr. 2009. "Tutorial on methods for interval-censored data and their implementation in R." *Statistical Modelling* 9(4):259–297.

Griffin, J. 2005. "INTCENS: Stata module to perform interval-censored survival analysis." *Statistical Software Components* .

Hartzell, Caroline and Matthew Hoddie. 2003. "Institutionalizing Peace: Power Sharing and Post-Civil War Conflict Management." *American Journal of Political Science* 47(2):318–332.

Heagerty, Patrick J. and Brenda F. Kurland. 2001. "Misspecified Maximum Likelihood Estimates and Generalised Linear Mixed Models." *Biometrika* 88(4):973–985.

Henschel, Volkmar, Christiane Heiss and Ulrich Mansmann. 2009. *intcox: Iterated Convex Minorant Algorithm for interval censored event data*. R package version 0.9.2.

Jewell, Nicholas P., Hina M. Malani and Eric Vittinghoff. 1994. "Nonparametric Estimation for a Form of Doubly Censored Data, With Application to Two Problems in AIDS." *Journal of the American Statistical Association* 89(425):7–18.

Kim, Dong K. 1997. "Regression Analysis of Interval-Censored Survival Data with Covariates Using Log-Linear Models." *Biometrics* 53(4):1274–1283.

Kim, Jong S. 2003. "Maximum Likelihood Estimation for the Proportional Hazards Model with Partly Interval-Censored Data." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 65(2):489–502.

King, Gary, James Honaker, Anne Joseph and Kenneth Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *The American Political Science Review* 95(1):49–69.

Krustev, Valentin L. 2006. "Interdependence and the Duration of Militarized Conflict." *Journal of Peace Research* 43(3):243–260.

Meier, Amalia S., Barbra A. Richardson and James P. Hughes. 2003. "Discrete Proportional Hazards Models for Mismeasured Outcomes." *Biometrics* 59(4):947–954.

Odell, Patricia M., Keaven M. Anderson and Ralph B. D'Agostino. 1992. "Maximum Likelihood Estimation for Interval-Censored Data Using a Weibull- Based Accelerated Failure Time Model." *Biometrics* 48(3):951–959.

Pan, Wei. 1999. "Extending the Iterative Convex Minorant Algorithm to the Cox Model for Interval-Censored Data." *Journal of Computational and Graphical Statistics* 8(1):109–120.

Pan, Wei. 2000. "A Multiple Imputation Approach to Cox Regression with Interval-Censored Data." *Biometrics* 56(1):199–203.

Pan, Wei. 2001. "A Multiple Imputation Approach to Regression Analysis for Doubly Censored Data with Application to AIDS Studies." *Biometrics* 57(4):1245–1250.

Pan, Wei and Rick Chappell. 2002. "Estimation in the Cox Proportional Hazards Model with Left-Truncated and Interval-Censored Data." *Biometrics* 58(1):64–70.

Peterson, David A. M., Lawrence J. Grossback, James A. Stimson and Amy Gangl. 2003. "Congressional Response to Mandate Elections." *American Journal of Political Science* 47(3):411–426.

R Development Core Team. 2008. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0. **URL:** *http://www.R-project.org*

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley & Sons.

Rubin, Donald B. 1996. "Multiple Imputation After 18+ Years." *Journal of the American Statistical Association* 91(434):473–489.

Satten, Glen A. 1996. "Rank-Based Inference in the Proportional Hazards Model for Interval Censored Data." *Biometrika* 83(2):355–370.

Satten, Glen A., Somnath Datta and John M. Williamson. 1998. "Inference Based on Imputed Failure Times for the Proportional Hazards Model with Interval-Censored Data." *Journal of the American Statistical Association* 93(441):318–327.

Senese, Paul D. and Stephen L. Quackenbush. 2003. "Sowing the Seeds of Conflict: The Effect of Dispute Settlements on Durations of Peace." *The Journal of Politics* 65(3):696–717.

Shipan, Charles R. and Megan L. Shannon. 2003. "Delaying Justice(s): A Duration Analysis of Supreme Court Confirmations." *American Journal of Political Science* 47(4):654–668.

Spriggs, James F., II and Thomas G. Hansford. 2001. "Explaining the Overruling of U.S. Supreme Court Precedent." *The Journal of Politics* 63(4):1091–1111.

StataCorp, LP. 2007. "Stata statistical software: release 10.0." *College Station, TX: Stata Corporation* .

Strom, Kaare and Stephen M. Swindle. 2002. "Strategic Parliamentary Dissolution." *The American Political Science Review* 96(3):575–591.

Sun, Jianguo, Qiming Liao and Marcello Pagano. 1999. "Regression Analysis of Doubly Censored Failure Time Data with Applications to AIDS Studies." *Biometrics* 55(3):909–914.

Therneau, Terry and original R port by Thomas Lumley. 2009. *survival: Survival analysis, including penalised likelihood.* R package version 2.35-7.
URL: *http://r-forge.r-project.org*

Turnbull, Bruce W. 1976. "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data." *Journal of the Royal Statistical Society. Series B (Methodological)* 38(3):290–295.

Vining, Richard L, Christopher Zorn and Susan Navarro Smelcer. 2006. "Judicial Tenure on the U.S. Supreme Court, 1790–1868: Frustration, Resignation, and Expiration on the Bench." *Studies in American Political Development* 20(02):198–210.

Volden, Craig. 2002. "The Politics of Competitive Federalism: A Race to the Bottom in Welfare Benefits?" *American Journal of Political Science* 46(2):352–363.

Wei, Zhang, Zhang Ying, Kathryn Chaloner and Jack T. Stapleton. 2009. "Imputation methods for doubly censored HIV data." *Journal of Statistical Computation Simulation* 79(10):1245 – 1257.

Wellner, Jon A. and Yihui Zhan. 1997. "A Hybrid Algorithm for Computation of the Nonparametric Maximum Likelihood Estimator From Censored Data." *Journal of the American Statistical Association* 92(439):945–959.

White, Halbert. 1982. "Maximum Likelihood Estimation of Misspecified Models." *Econometrica* 50(1):1–25.

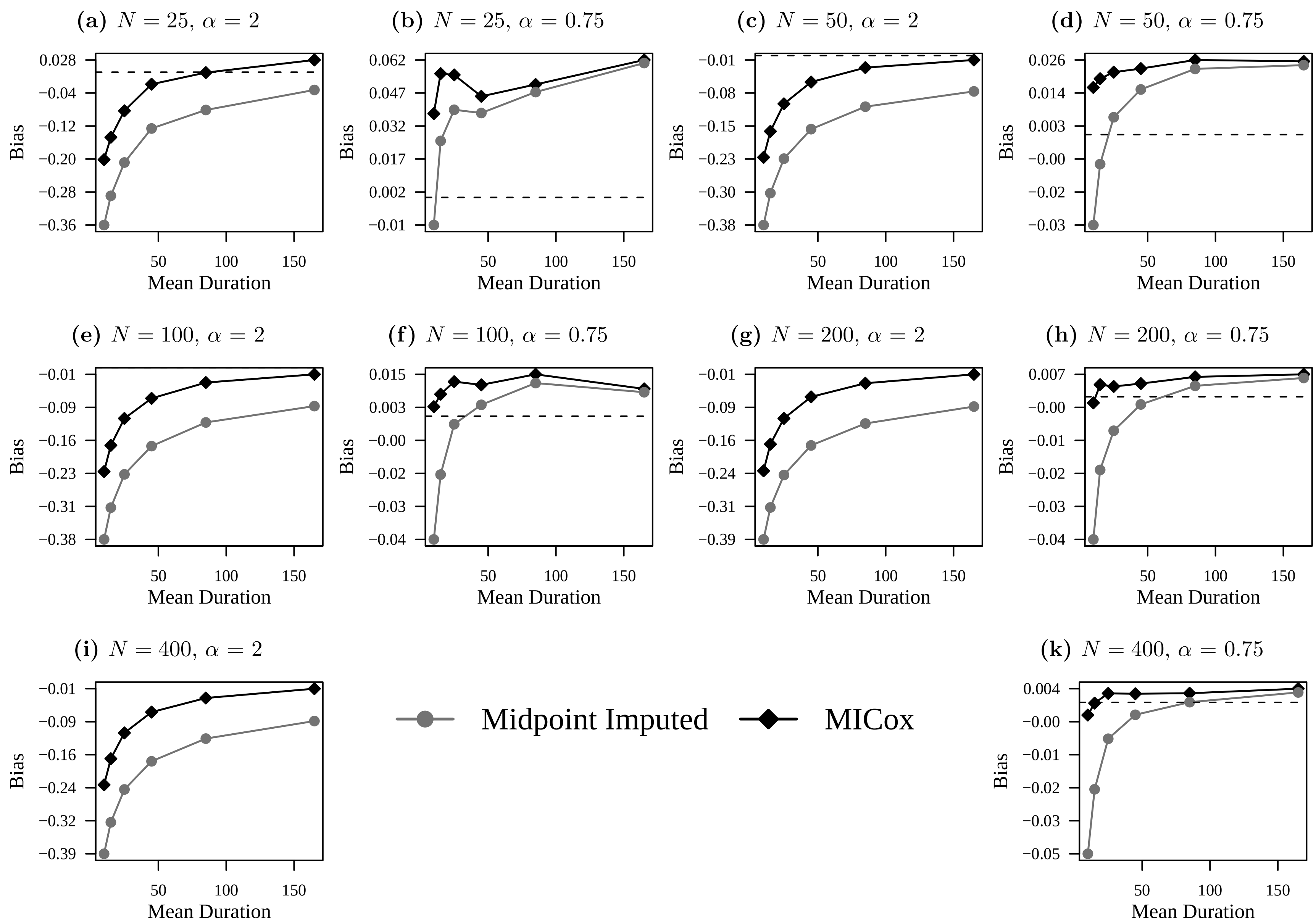

Figure 1: Bias in the Monte Carlo Study. Estimate given is the difference between the average value of the estimated regression coefficient and its true value of one. Each point is derived from the 5000 replications specific to the sample size and α given in the plot title, and the mean duration given on the x-axis of the plot.

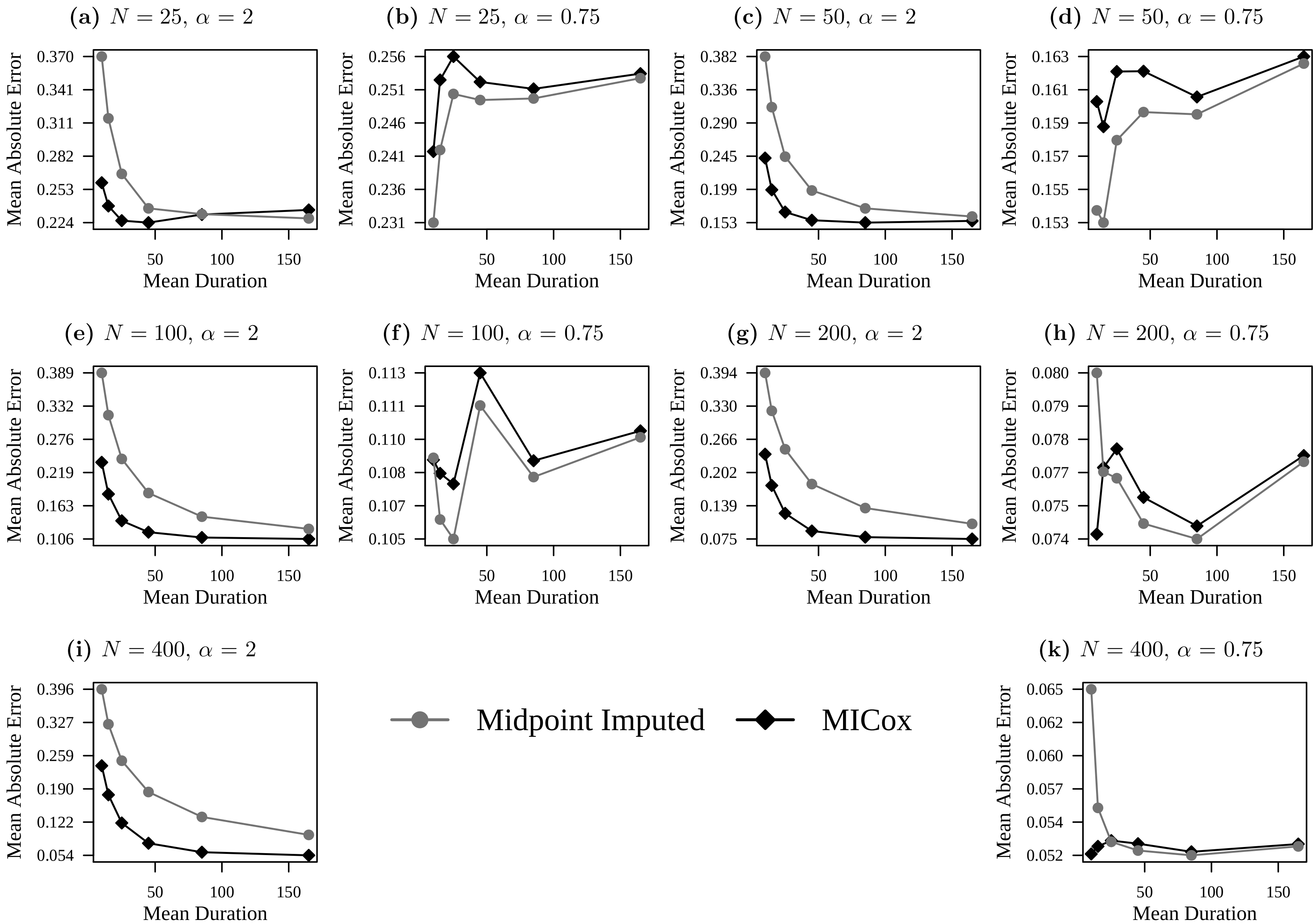

Figure 2: Mean absolute error in the Monte Carlo Study. Estimate given is the average absolute difference between the average value of the estimated regression coefficient and its true value of one. Each point is derived from the 5000 replications specific to the sample size and α given in the plot title, and the mean duration given on the x-axis of the plot.

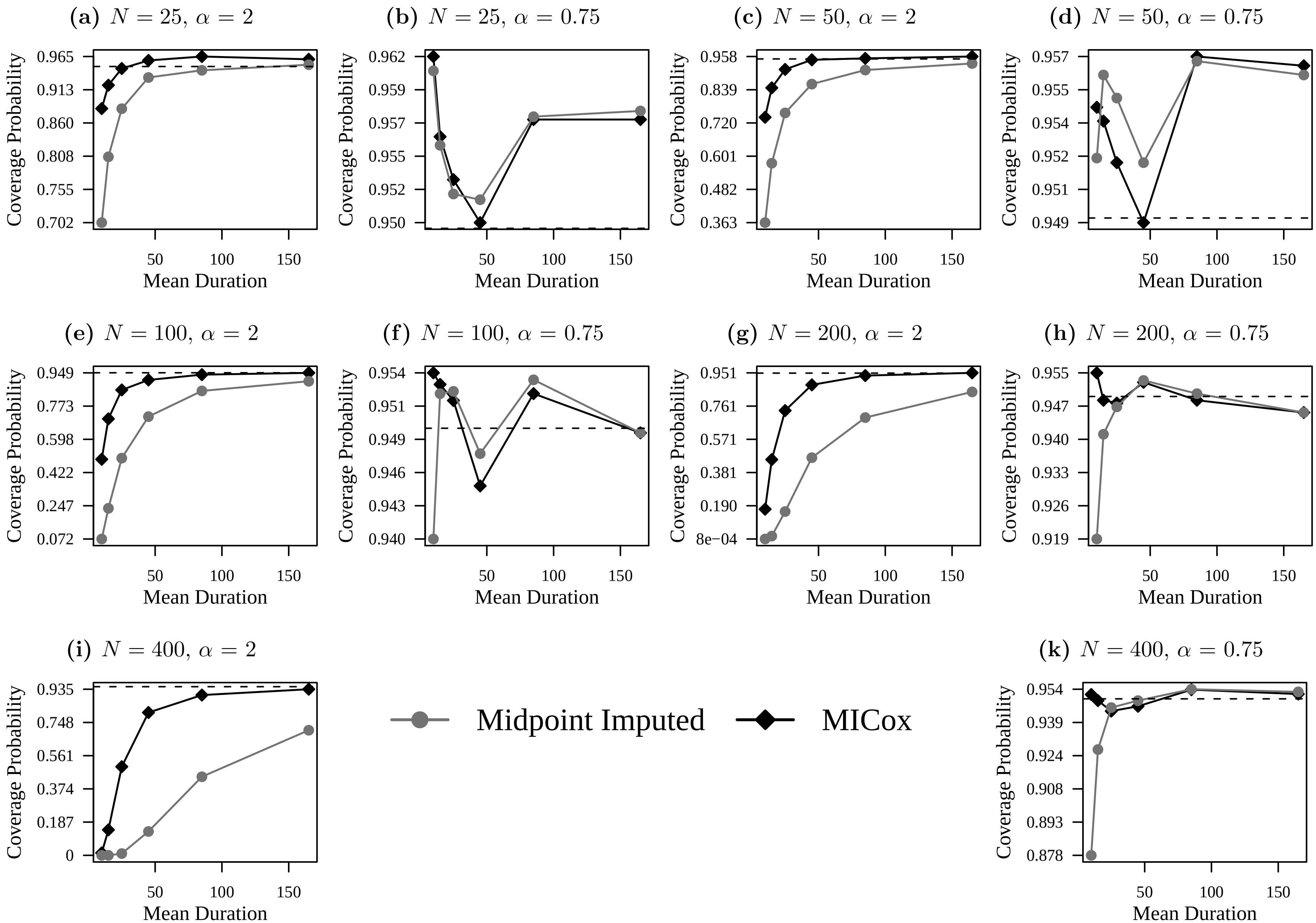

Figure 3: Coverage probability of the nominal 95% confidence interval in the Monte Carlo Study. Estimate given is the proportion of replications in which the 95% confidence interval contained one. Each point is derived from the 5000 replications specific to the sample size and α given in the plot title, and the mean duration given on the x-axis of the plot.

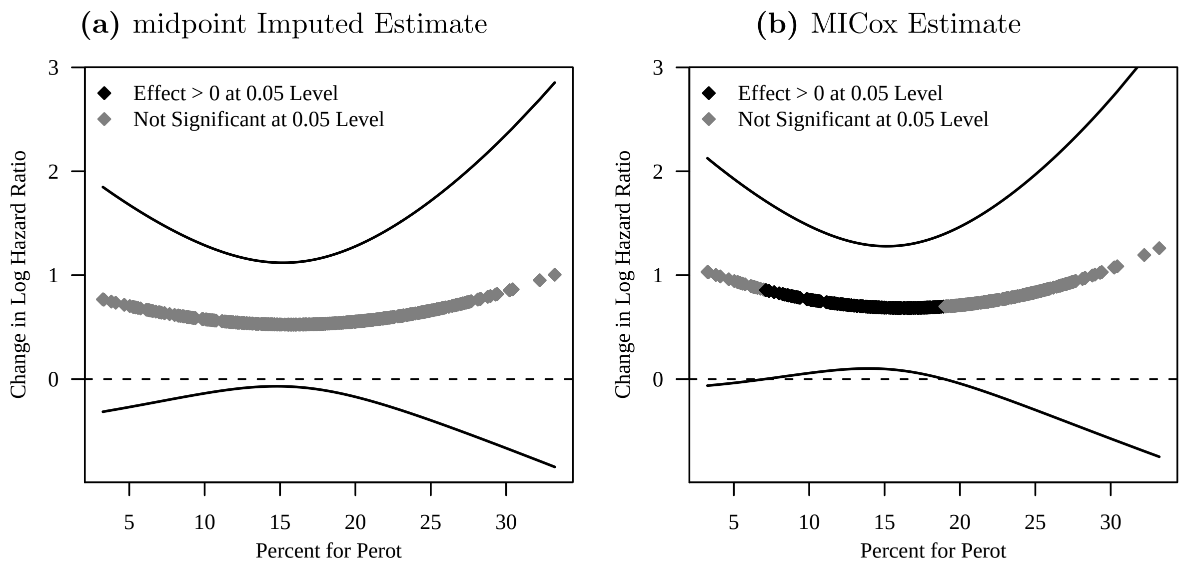

Figure 4: Effect of the percent voting for Perot on the log hazard ratio (e.g. the Cox regression coefficient) for the time to position-taking on NAFTA. The solid lines indicate two-sided 90% confidence limits. The vote values for points in the plots are the 433 actual values of the district-level vote for Perot from the sample in Box-Steffensmeier, Arnold and Zorn (1997).

	Model (3)			*Model (4)*		
	Mid	*Censored*	*% Δ*	*Mid*	*Censored*	*% Δ*
Divided Control	-0.65*	-0.73*	12	–	–	–
	(-2.83)	(-2.13)	-25			
President-Majority Party Distance	–	–	–	-0.78*	-0.92*	18
				(-1.86)	(-1.79)	-3.6
Age	0.02	0.027	33	0.02	0.028	38
	(1.33)	(1.22)	-8.6	(1.33)	(1.28)	-3.6
Current Senator	2.6*	3*	16	2.6*	2.9*	14
	(5.45)	(4.39)	-19	(5.45)	(4.34)	-20
Previous Senator	-0.36	-0.25	-31	-0.32	-0.21	-34
	(-0.692)	(-0.599)	-13	(-0.604)	(-0.504)	-17
State Court Experience	0.64*	0.75*	17	0.65*	0.76*	17
	(2.46)	(2.67)	8.6	(2.5)	(2.71)	8.2
Federal District Court Experience	0.83*	0.78	-6.3+	0.84*	0.81	-3.9+
	(1.84)	(1.56)	-15	(1.87)	(1.62)	-13
Federal Appellate Court Experience	-0.2	-0.22	12	-0.23	-0.25	6.5
	(-0.769)	(-0.782)	1.6	(-0.885)	(-0.857)	-3.2
Chief Justice	-0.65	-0.37	-43	-0.62	-0.35	-44
	(-1.48)	(-1.02)	-31	(-1.41)	(-0.955)	-32
Critical Nomination	-1.4*	-1.6*	13	-1.4*	-1.5*	10
	(-1.96)	(-3.08)	57	(-1.92)	(-2.94)	54
Critical Nomination X Divided Control	1.5*	1.8*	23	1.2	1.5*	22+
	(1.65)	(2.22)	34	(1.42)	(1.91)	34
Time Remaining in Session	-0.003	-0.0029	-3.4	-0.002	-0.0028	38
	(-3)	(-2.19)	-27	(-2)	(-2.09)	4.7
Last Year of Pres. Term	-0.4	-0.34	-14	-0.37	-0.3	-18
	(-0.727)	(-0.938)	29	(-0.673)	(-0.836)	24
Unelected President	0.52	0.73*	41+	0.54	0.7	30
	(1.13)	(1.67)	47	(1.15)	(1.60)	39

Table 1: Replication of the analysis from table [2] in Shipan and Shannon (2003). Change in log hazard ratios reported. Z-statistics appear in parentheses. *Mid* = midpoint imputed, *Censored* = interval censored estimation. * - statistically significant at the 0.05 level (one-tailed). N= 87 in all analyses.

	Mid	*Censored*	*% Δ*
Union Membership	3.21*	3.89*	21.31
	(2.697)	(3.22)	19.51
Perot Vote, Percent	-4.91	-6.61	34.6
	(-1.15)	(-1.53)	32.69
Perot Vote, Percent Squared	15.64	20.2*	29.23+
	(1.334)	(1.68)	25.84
Mexican Border	1.84*	2.14*	16.49
	(5.75)	(6.55)	13.92
Household Income	0.01	0.019	90.22
	(0.111)	(0.209)	87.81
Corporate Contributions	-1.44*	-1.47*	1.76
	(-2.717)	(-2.78)	2.31
Labor Contributions	1.09*	1.13*	3.68
	(2.18)	(2.22)	1.99
NAFTA Committee	0.04	0.0116	-71.11
	(0.364)	(0.106)	-70.98
Republican Leadership	0.56*	0.551*	-1.59
	(2.154)	(2.14)	-0.84
Democratic Leadership	0.08	0.103	29.07
	(0.348)	(0.44)	26.59
Ideology X Union Membership	-4.39*	-5.36*	22.11
	(-2.466)	(-2.98)	20.93
Ideology X Household Income	0.16	0.148	-7.56
	(1.231)	(1.17)	-4.88

Table 2: Replication of the analysis from table [2] in Box-Steffensmeier, Arnold and Zorn (1997). Change in log hazard ratios reported. Z-statistics appear in parentheses. *Mid* = midpoint imputed, *Censored* = interval censored estimation. * - statistically significant at the 0.05 level (one-tailed). N= 433 in all analyses.

	Mid	*Censored*	% Δ
Coup/revolution	0.32*	0.309*	-3.4
	(-5.36)	(-5.07)	-5.35
Eastern Europe	0.33*	0.321*	-2.58
	(-4.21)	(-3.96)	-5.91
Non contiguous	0.684	0.681	-0.506
	(-1.62)	(-1.55)	-4.55
Sons of the Soil	3.1*	3.2*	3.18
	(3.86)	(3.74)	-3.07
Contraband	2.56*	2.66*	3.66
	(2.76)	(2.7)	-2.22

Table 3: Replication of model (1) from table [2] in Fearon (2004). Multiplicative effect of the independent variable on the expected duration is reported. Z-statistics appear in parentheses. *Mid* = midpoint imputed, *Censored* = interval censored estimation. * - statistically significant at the 0.05 level (one-tailed). N = 128 in all analyses.

	Cox			*Weibull*		
	Mid	*Censored*	% Δ	*Mid*	*Censored*	% Δ
Power-sharing Institutions	-0.79*	-0.3	-62.19+	-0.76*	-0.81*	6.66
	(-2.08)	(-1.4)	-33.72	(-2.05)	(-2.3)	9.72
Previous Regime Type	-0.71	-0.26	-63.07	-0.6	-0.65	8.12
	(-0.96)	(-0.54)	-43.79	(-0.83)	(-0.91)	8.93
Conflict Duration	-0.74	-0.075	-89.92	-0.73	-0.79	8.74
	(-1.19)	(-0.19)	-84.48	(-1.24)	(-1.3)	6.33
Conflict Intensity	0.79*	0.14	-82.55+	0.81*	0.87*	6.92
	(2.93)	(1.0)	-65.2	(3.38)	(3.6)	7.53
Third Party Enforcer	-1.86*	-0.36	-80.73+	-1.77*	-2*	10.21
	(-2.38)	(-0.76)	-68.3	(-2.53)	(-2.6)	2.01
International System Structure	-0.26	1.7*	552.76+	0.06	-0.11	81.11
	(-0.4)	(3.2)	703.44	(0.1)	(0.18)	73.96
Conflict Issue	1.55*	0.46	-70.19+	1.63*	1.7*	3.62
	(1.89)	(1.2)	-36.78	(2.06)	(2.2)	6.95

Table 4: Replication of the analysis from table [2] in Hartzell and Hoddie (2003). Change in log hazard ratios reported. Z-statistics appear in parentheses. *Mid* = midpoint imputed, *Censored* = interval censored estimation. * - statistically significant at the 0.05 level (one-tailed). N= 38 in all analyses.